DISCOVERY OF CANDIDATE GENES FOR STALLION FERTILITY FROM THE HORSE Y CHROMOSOME

A Dissertation

by

NANDINA PARIA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2009

Major Subject: Biomedical Sciences

DISCOVERY OF CANDIDATE GENES FOR STALLION FERTILITY FROM THE HORSE Y CHROMOSOME

A Dissertation

by

NANDINA PARIA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Terje Raudsepp
Committee Members,	Bhanu P. Chowdhary
	William J. Murphy
	Paul B. Samollow
	Dickson D. Varner
Head of Department,	Evelyn Tiffany-Castiglioni

August 2009

Major Subject: Biomedical Sciences

ABSTRACT

Discovery of Candidate Genes for Stallion Fertility from the Horse Y Chromosome.

(August 2009)

Nandina Paria, B.S., University of Calcutta;

M.S., University of Calcutta

Chair of Advisory Committee: Dr. Terje Raudsepp

The genetic component of mammalian male fertility is complex and involves thousands of genes. The majority of these genes are distributed on autosomes and the X chromosome, while a small number are located on the Y chromosome. Human and mouse studies demonstrate that the most critical Y-linked male fertility genes are present in multiple copies, show testis-specific expression and are different between species.

In the equine industry, where stallions are selected according to pedigrees and athletic abilities but not for reproductive performance, reduced fertility of many breeding stallions is a recognized problem. Therefore, the aim of the present research was to acquire comprehensive information about the organization of the horse Y chromosome (ECAY), identify Y-linked genes and investigate potential candidate genes regulating stallion fertility.

To achieve theses goals, a direct cDNA (complementary DNA) selection procedure was used to isolate Y-linked genes from horse testes and 29 Y-specific genes were identified. All 29 genes were mapped to ECAY and their sequences were used to

further expand the existing map. Copy number analysis identified 15 multicopy genes of which 9 were novel transcripts. Gene expression analysis on a panel of selected body tissues showed that some ECAY genes are expressed exclusively in testes while others show ubiquitous or intermediate expression. Quantitative Real-Time PCR using primers for 9 testis-specific multicopy genes revealed 5 genes with statistically significant differential expression in testis of normal fertile stallions and stallions with impaired fertility. Gene copy number analysis showed that the average copy number of 4 such genes was decreased in subfertile/infertile stallions compared to normal animals.

Taken together, this research generated the first comprehensive physical gene map for the horse Y chromosome and identified a number of candidate genes for stallion fertility. The findings essentially expand our knowledge about Y chromosome genes in horses, open a new avenue for investigating the potential role of ECAY genes in stallion fertility which contribute to the development of molecular tools for the assessment of fertility in stallions.

DEDICATION

This work is dedicated to my loving parents, Mrs. Kalpana Paria and Dr. Nandadulal Paria. This research would not have been possible without their blessings, encouragement and constant support.

ACKNOWLEDGEMENTS

The journey of three and half years of my graduate program has been quite eventful and memorable with the help and co-operation of a lot of people to whom I'm really grateful from the bottom of my heart.

First and foremost, I would like to express my sincere gratitude to my major advisor Dr. Terje Raudsepp for her valuable guidance, suggestion and assistance throughout my entire project. I cannot remember a single incident when I required her help and she was not available!!! Thanks a lot, Terje, whatever I learnt from you, is going to be a big asset for my future.

I'm grateful to all of my committee members for their immense help and co-operation during my research. I'm extremely thankful to Dr. Bhanu Chowdhary, who was more like my unofficial coadvisor. His valuable guidance always kept me on track and on top of things. He played a very significant role in my graduate career and I will always be grateful to him for his constant support. I would like to gratefully acknowledge Dr. William Murphy for patiently teaching me some of the experiments, helping me to troubleshoot and of course for his valuable suggestions. I would like to thank Dr. Paul Samollow for his insightful discussions and suggestions to improve my research. Many thanks go to Dr. Dickson Varner for his immense help in providing tissue samples for my study and valuable suggestions during my research.

I'm especially thankful to all of my past and present lab members who helped me in this project as well as made my stay in this lab a wonderful one! I am especially

grateful to Dr. Ashley Seabury for being a great office-mate, mentor and friend. She helped me in all possible ways whenever I encountered any difficulty. Many thanks go to Dr. Keith Durkin for his endless sequencing and helping me with tissue sample collection. I gratefully acknowledge Dr. Sankar Chaki for technical assistance and his jovial jestures that kept me smiling, Dr. Pranab Das for helping me with tissue collection and sequencing, Dr. Dhruva Mishra and Dr. Monica Vishnoi for being great labmates. I am also thankful to my fellow graduate students Priyanka Kachroo, Jana Caldwell and Jixin Wang who helped me either by providing tissue samples for my experiments or teaching me how to use new software and for being great co-workers! I would also like to thank my past lab members Dr. Christopher Seabury, Dr. Eun-Joon Lee, Dr. Deeann Wallis and Dr. Candice Brinkmeyer-Langford for their assistance during the early phase of my tenure.

I would also like to thank Dr. Charle Love and Dr. Caroline Arnold for providing tissue samples, Dr. Loren Skow for letting me use the TAMU equine BAC library. My sincere thanks go to Alison Pearks Wilkerson who patiently taught me some of my experiments and she always had answers for my endless queries with a smiling face.

My sincere gratitude goes to Dr. Brian Williams and the members of my support group at SCS who have been a major help during my entire process of disseratation writing. Thank you all.

I am especially thankful to my friends Sucharita Bose, Soumya Ganguly, Rangan Gupta and Sharmila Ghosh whose support means a lot to me. Their love and friendship helped me to overcome any hurdle that I faced during my entire stay in USA.

And finally I would like to acknowledge my parents, grandparents and all other family members who are thousands of miles away from me but still were a constant source of love, encouragement and blessings for me. They gave me the courage and strength to accomplish the goal that I came here for. Love you all.

NOMENCLATURE

BAC	Bacterial Artificial Chromosome
BES	BAC End Sequence
cDNA	Complementary DNA
ECA	Equus caballus, horse
ECAY	Horse Y chromosome
EST	Expressed Sequence Tag
FCA	Felis catus, cat
HSA	Homo sapiens, human
MMU	Mus musculus, mouse
MSY	Male-specific region of Y chromosome
PCR	Polymerase Chain Reaction
qRT-PCR	Quantitative Real Time PCR
RACE	Rapid Amplification of cDNA Ends
RH	Radiation Hybrid
RT-PCR	Reverse Transcriptase PCR
SNP	Single Nucleotide Polymorphism
STS	Sequence Tagged Site

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

GENERAL INTRODUCTION

OVERVIEW OF THE GENETICS OF MALE FERTILITY IN MAMMALS

Male fertility in mammals is a complex trait and is governed by a combination of multiple environmental and genetic factors. Though only limited information is available about the latter, it has been proposed that almost 20% of the genes in mammalian genomes (~5,000 genes) are involved in different aspects of male fertility - sex determination, sexual differentiation and testicular development, spermatogenesis, sperm function and sperm-oocyte interactions preceding fertilization (Hargreave 2000; Matzuk and Lamb 2002; Carrell 2007; Krausz and Giachini 2007). Male fertility is the result of a well orchestrated and spatio-temporally regulated interaction of these genes, though our current knowledge about the role of specific genes at different stages of male development is limited. This also implies that very little is known about the underlying molecular causes of male infertility. To date only a few hundreds male fertility genes have been identified using mainly transgenic, knockout, or mutant rodent models (Matzuk and Lamb 2002; Carrell 2008; Matzuk and Lamb 2008). In the following sections a brief overview of current knowledge about the genetic regulation of the most critical steps in male fertility is presented.

This dissertation follows the style of Cytogenetic and Genome Research.

Male sex determination

Male sex determination in mammals is a complex process that involves a cascade of molecular events and interactions among many genes. In eutherian mammals the onset of maleness is governed by a single gene - *SRY* (Sex-determining Region on Y) that encodes a high mobility group (HMG) family transcription factor protein. In mouse *Sry* transcripts are expressed for a brief period during early stages of gonadal development (between 10.5-12.5 days postcoitum) and the ability of *SRY* to induce testis development is limited to a time window of only 6 hours after the normal onset of its expression in XY gonads. This is an indication that *SRY* mainly initiates testis development rather than maintains testis differentiation (Hiramatsu et al. 2009; Sekido and Lovell-Badge 2009). Several hypotheses have been proposed regarding *SRY* function in sex determination. For example, *SRY* might antagonize a repressor of male development or might initiate transcription of one or more genes that have important roles in male development (Wilhelm et al. 2007a). However, in order to understand *SRY* functional pathway(s), knowledge of *SRY* target genes is required. With recent progress in deciphering the genetic component of mammalian sex determination, several other genes with important roles in sex determination and male development have been discovered. Among these are SRY-box 9 (*SOX9),* steroidogenic factor (*SF1,* now known as nuclear receptor subfamily 5, group A, member 1, *NR5A1*), GATA-binding protein 4 (*GATA4*), fibroblast growth factor 9 (*FGF9*), wingless-type MMTV integration site family, member 4 *(WNT4)*, R-Spondin1 (*RSPO1*) and many more. Rodent studies provide compelling evidence that the only downstream target of *SRY* is *SOX9* which in turn

activates the rest of the male pathway genes, drives Sertoli cell formation and hence, testis differentiation (Sekido et al. 2004; Wilhelm et al. 2005; Sekido and Lovell-Badge 2008). Murine *Sry* binds to multiple elements within a *Sox9* gonad-specific enhancer and upregulates *Sox9.* Although *Sry* initiates maleness in the embryo, *Sox9* maintains the later part of male development. While *Sry* expression is transient, *Sox9* expression is maintained in Sertoli cells throughout life (Sekido and Lovell-Badge 2009). Furthermore, *SRY* upregulates *SOX9* expression rather than initiates it. It is proposed that *SOX9* transcriptional regulation consists of at least three phases: i) *SRY*-independent and probably *SF1*-dependent initiation, ii) *SRY*-dependent upregulation, and iii) *SRY*-independent maintenance (Sekido et al. 2004). This also explains how gain-of-function mutations of *SOX9* in humans and mice can initiate male development in female (XX) embryos in the absense of *SRY* (Polanco and Koopman 2007; Wilhelm et al. 2007b). Another key-role gene in the self-reinforcing pathway of male sex determination is *SF1* which initiates *SOX9* transcription and thereafter, cooperatively with *SRY,* upregulates *SOX9* expression. A positive feed-back loop between *SF1* and *SOX9* is also needed to maintain *SF1* transcription. After *SRY* expression has ceased, *SOX9* binds to its enhancer along with *SF1* to maintain *SOX9* expression in the absence of *SRY*. Therefore, mammalian sex determination involves the synergistic action of *SRY* and *SF1* on a specific *SOX9* enhancer to upregulate and maintain *SOX9* expression (Sekido and Lovell-Badge 2008). Additionally, *SOX9* binds to its own enhancer and self-regulates its expression. These early events in male sex determination involve mainly three critical genes – *SRY*, *SF1*, *SOX9* and focus on the initiation and maintenance of *SOX9*

expression. *SOX9*, in turn, is probably the pivotal and distinguishing factor that influences the expression of genes that define Sertoli cell phenotype and function, such as anti-Mullerian hormone (*AMH)* and prostaglandin D2 synthase (*PTGDS)*. A positive feed-forward loop between *SOX9* and *FGF9* activates the *FGF9-FGFR2* signalling pathway which antagonizes the activity of *RSPO1-WNT4* and suppresses ovarian development (Kim et al. 2006; Wilhelm 2007). *FGF9*, in turn, is indispensable to maintain *SOX9* expression in Sertoli precursor cells. *DAX1* acts as an anti-testis gene by antagonizing the function of *SRY* in mammalian sex determination pathways (Swain et al. 1998). While *SOX9* activates genes critical for male development, it also represses the activity of genes typical to bipotential cell precursors and follicle cells. Once *SOX9* activity has reached a critical threshold, it represses *SRY* activity, as well as the activity of typical female pathway genes such as *WNT4*, *RSPO1, DAX1* and *FOXL2* (Nef et al. 2003; Sekido and Lovell-Badge 2008). In summary, male sex determination in mammals is a complex process which is sensitive to gene dosages, involves synergistic interactions of many genes forming a network of positive and negative feedback or feed forward loops and keeps a fine balance between male and female developmental pathways (Sekido and Lovell-Badge 2008). However, it should be noted that most current knowledge about male sex determination is based on mouse models and might not reflect the situation in other mammalian species. A schematic representation of *SRY-SOX9* interplay in mammalian sex determination pathways is shown in Fig. 1.

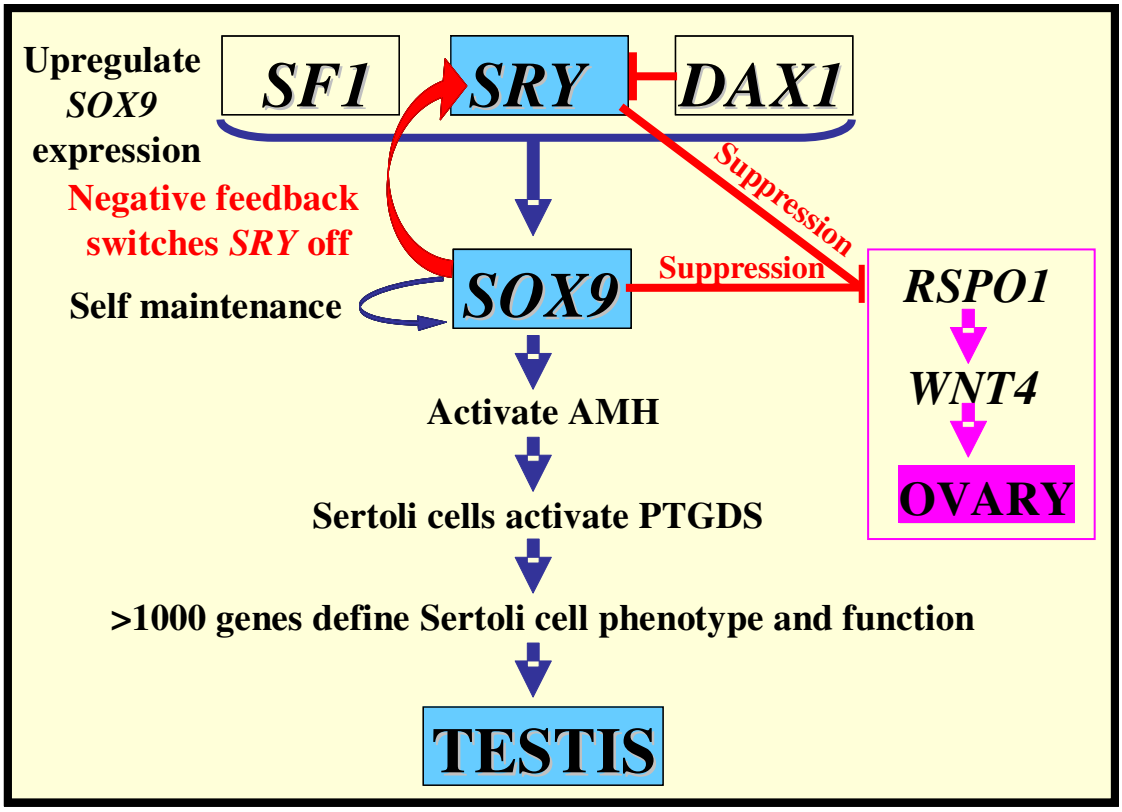

Figure 1: Schematic representation of *SOX9-SRY* interactions at the onset of mammalian sex determination (Wilhelm 2007, Wilhelm et al. 2007b, Sekido and Lovell-Badge 2008).

Male sexual differentiation

Once the sex of the embryo is determined as male, the bipotential gonad differentiates along the male pathway into a testis which, in turn, dictates the development of male secondary sex organs and the differentiation of Wolffian ducts into the male genital tract, *viz.,* epididymis, *vas deferens* and seminal vesicles (Matzuk and Lamb 2008). All these events, especially the processes that lead to the proper

development, growth, and descent of testes, are critical for spermatogenesis and male fertility.

The complex events of male sexual differentiation are sensitive to environmental influences (e.g., toxins, temperature) and are regulated by many genes coding for various growth and transcription factors, hormones and signaling molecules (Matzuk and Lamb 2008). As already mentioned, testis differentiation is induced by the expression of S*RY* and *SOX9* in a subset of somatic cells, known as Sertoli cell precursors that develop into Sertoli cells. Sertoli cells, in turn, orchestrate the differentiation of all other cell types in males by secreting appropriate levels of signaling molecules (Wilhelm et al. 2007b). This leads to the formation of testis cords (the precursors of adult spermatogenic tubules), differentiation of Leydig cells, vascular cells and other interstitial cells such as fibroblasts, mast cells, macrophages and lymphocytes. Leydig cells are important for secreting the hormones that play vital roles in establishing and maintaining the secondary male sex characteristics (Wilhelm et al. 2007b). Increased proliferation and migration of peritubular myoid cells, Leydig cells and endothelial cells from adjacent mesonephros are responsible for the increase of testis size (Schmahl et al. 2000).

An important event in male differentiation is proper descent of the testes. Impaired testicular descent, cryptorchidism, exposes the undescended testis for a prolonged time to increased temperature resulting in compromised spermatogenesis and reduced fertility or infertility. Cryptorchdism as a possible cause of impaired spermatogenesis was first described over 80 years ago (Crew 1922) and is one of the best-characterized risk factors for reduced male fertility (Foresta et al. 2008). The

condition is a more common problem (2-12%) in humans, pigs, dogs, cats and horses compared to cattle and sheep (<1%) (Amann and Veeramachaneni 2007). Despite this, very little is known about the genetic causes of cryptorchidism. The Mouse Genome Informatics (MGI: http://www.informatics.jax.org/[1]) database describes about 44 genes and markers that are associated with the murine cryptorchid phenotype. Eighteen of these genes are also believed to be involved in unilateral or bilateral cryptorchidism in humans (Matzuk and Lamb 2008), though causative mutations have been found in only a few genes and a few individuals. For example, a single nucleotide mutation in insulin-like factor 3 (*INSL3*) is associated with some, but not all, cases of human cryptorchidism (Feng et al. 2004; Agoulnik and Feng 2007). *INSL3* is required for gradual expansion of the gubernaculum which, in turn, provides an anchor for normal abdominal translocation of the testis. Other critical candidate genes for cyptorchidism are *INSL3* receptor (*LGR8,* now known as *RXFP2*), homeo box A10 (*HOXA10*) and estrogen receptor 1 *(ESR1*) (Matzuk and Lamb 2008). Over-expression of the aromatase gene (*CYP19A1*) can cause cryptorchidism by raising intra-testicular estradiol levels, introducing hormonal imbalance in the system (Klonisch et al. 2004). Additionally, environmental factors that interfere with the endocrine regulation of testicular descent might also contribute to the etiology of cryptorchidism (Foresta et al. 2008) and complicate the search for genetic causes.

Compared to the well studied physiology and biochemical pathways of signaling molecules leading to the differentiation of primordial germ cells, not much is known about the genetic basis of male sexual development and differentiation in mammals.

There is a non-exhaustive list of about 90 candidate genes that might be associated with various stages of male differentiation, while single nucleotide polymorphisms (SNPs) or mutations have been described for only seven genes, *viz., ATF3, CFTR, ESR1, ESR2, FGFR2, INSL3, SF1* (Matzuk and Lamb 2008). Similarly to male sex determination, most of the genetic studies on male sexual differentiation are based on rodent models and only a few have searched for candidate genes and mutations in humans or domestic animals (Matzuk and Lamb 2002; Wilhelm et al. 2007b; Barthold 2008; Foresta et al. 2008).

Spermatogenesis

Mammalian spermatogenesis entails a sequence of events that are initiated in the testis, proceed in the epididymis, and culminate with sperm capacitation in the female reproductive tract. For successful sperm, these events are followed by acrosomal exocytosis, penetration of the oocyte zona pellucida, and fertilization (Luk et al. 2006; Varner and Johnson 2007; Marengo 2008). Spermatogenesis involves mitotic and meiotic cell divisions and both nuclear and cytoplasmic reorganizations, thus, many proteins are transiently or temporally expressed in a particular stage of the spermatogenic cycle (Cheng and Mruk 2002). This implies that spermatogenesis is under spatio-temporal control of many genes. A variety of approaches have identified over 1000 genes that are specifically expressed in male germline and become active in testis exemplifying the complexity of the process. This also indicates that mutations in thousands of different genes could cause male infertility (Cooke and Saunders 2002;

Nayernia et al. 2003; Chenoweth 2005; Carrell 2007). Targeting specific genes that regulate particular processes during spermatogenesis, however, has had limited success, and has been carried out mainly in model species like mice and rats (Matzuk and Lamb 2002; Nayernia et al. 2003). In contrast, very little is known about the genes controlling spermatogenesis and male fertility in other mammalian species, including humans and horses. To illustrate the complex genetics underlying spermatogenesis, a summary of events at different phases of this process is described below.

i) Mitotic and meiotic germ cell divisions

Male germ cell proliferation begins in embryogenesis and with the exception of a short prenatal-prepubertal period, spermatogonial stem cells proliferate throughout life. Spermatogenesis, the development of mature sperm from spermatogonia, starts at puberty under the influence of the hormone testosterone. Spermatogonial cells differentiate into diploid primary spermatocytes which then undergo meiosis. Meiosis is a type of cell division unique to germ cells and is needed for the production of haploid gametes for sexual reproduction. Synapsis of homologous chromosomes during the prophase of the first meiotic division is necessary for proper segregation of homologous chromosomes into haploid spermatids, while meiotic recombination increases genetic variation by exchanging alleles between paternal and maternal homologs. Recombination and faultless segregation of homologous chromosomes are, thus, among the most critical events during spermatogenesis and any mistake results in aberrant

sperm. The most extreme consequence of meiotic errors is the production of semen without sperm (azoospermia) causing complete infertility.

Mitotic and meiotic cell divisions are controlled by hundreds of genes. Recent studies of transcriptional regulatory network of the mitotic cell cycle in normal human fibroblasts identified 480 periodically expressed genes (Bar-Joseph et al. 2008) and it could be expected that meiosis is regulated by even more genes. Among the approximately 115 genes associated with various types of sperm defects (O'Bryan and de Kretser 2006; Matzuk and Lamb 2008), several are critical for meiotic recombination and chromosome movements during meiosis. To mention a few, male mice lacking both copies of synaptonemal complex protein 3 (*SYCP3*) produce no sperm (Hughes 2008), mutation in chromosomal passenger protein aurora kinase C gene (AURKC) which is needed for normal centromere function, leads to the production of large-headed polyploid multi-flagellar spermatozoa causing infertility in human males (Dieterich et al. 2009) and mutations in recombination genes such as *SPO11*, *DMC1*, *MSH4* and *MSH5* cause complete arrest of germ cells at zygonema of meiosis I prophase and result in azoospermia both in men and mice (Gonsalves et al. 2004). Normally, aberrant germ cells are arrested at cell cycle checkpoints and subjected to DNA repair or, if damaged beyond repair, eliminated by programmed cell death also known as apoptosis (Gonsalves et al. 2004). Consequently, mutations in genes controlling cell cycle checkpoints, DNA repair mechanisms or apoptosis contribute to the production of various types of defective sperm (Matzuk and Lamb 2008).

ii) Spermiogenesis

The final phase of sperm development within testis is spermiogenesis. During this stage the haploid round spermatids undergo dramatic transformation and are finally released during the process called spermiation into the lumen of seminiferous tubules as spermatozoa (Beardsley and O'Donnell 2003; Varner and Johnson 2007). Terminal differentiation of spermatids involves four major events: i) acrosome biogenesis, ii) formation of sperm tail or flagellum, iii) chromatin condensation, and iv) removal of most of the sperm cytoplasm. These complex processes are critical for the development of fully functional sperm and at this time the sperm are most vulnerable to both structural and genetic defects (Varner and Johnson 2007).

The acrosome is a large secretory vesicle located in the sperm's head between the nucleus and the plasma membrane and is essential for penetration of the sperm through the oocyte's zona pellucida and fertilization (Mayorga et al. 2007; Varner and Johnson 2007; Zhao et al. 2007; Zhao et al. 2008). The absence of the acrosome, also known as globozoospermia, is a genetic disorder of male infertility and is characterized by the presence of 100% round-headed sperm lacking an acrosome (Dam et al. 2007). Studies in humans and mice agree that defects in acrosome biogenesis have a genetic basis, though very little is known about candidate genes and mutations. MGI (http://www.informatics.jax.org/[1]) describes about 40 murine genes and phenotypes with an abnormal or missing acrosome of which some though not all might be associated with similar conditions in humans or other mammals. For example, homozygous deletion of the casein kinase 2 (*Csnk2a2)* causes globozoospermia in mouse (Rocha and Affara

2000; Truong et al. 2003) but in spite of expectations, no mutations in *CSNK2A2* have been detected in globozoospermic human patients (Pirrello et al. 2005).

The development of the sperm tail is another key-process which directly affects sperm motility and involves formation of axoneme, outer dense fibres, fibrous sheath and mitochondrial sheath (Varner and Johnson 2007; Matzuk and Lamb 2008). Mouse models showing impaired sperm motility, also known as asthenozoospermia, provide some clues concerning potential gene defects in humans. Among proposed candidate genes for human asthenozoospermia, several encode known flagellar proteins such as axoneme associated proteins *SPAG6* and *SPAG16,* A kinase (PRKA) anchor protein 4 (*AKAP4*) and genes encoding proteins in axonemal dynein cluster (*DNAI1*, *DNAH5* and *DNAH11* (O'Bryan and de Kretser 2006; Matzuk and Lamb 2008). Sperm motility can be affected also by deletions and mutations in mitochondrial genes such as *MTCYB* and *MTATP6* (Feng et al. 2008), genes governing sperm transit from epididymis, such as homeobox 5 (*Rhox5*) (Shanker et al. 2008), sperm without mobility 2 (*SWM2*) (Lessard et al. 2007) and genes regulating intracellular calcium channels and potassium currents in sperm, such as the *CATSPER* gene family (Matzuk and Lamb 2008).

iii) Epididymal transit and sperm maturation

Though sperm chromatin condensation starts during spermiogenesis, most of it takes place during sperm maturation in the epididymis. This process involves extensive cellular remodeling and results in packing the haploid genome of sperm into a compact transcriptionally silent structure where most histones are replaced with protamines

(Braun 2001; Varner and Johnson 2007). After maturation, sperm chromatin packing level exceeds that of a diploid somatic cell by about six fold (Braun 2001). Sperm chromatin remodeling has several unique features and requires expression of many male-specific genes of which the most important and unique are protamines (Braun 2001). Sperm maturation is a systematic and gradual process. First, somatic histones are partly replaced by testis-specific histones which then are replaced by transitional nuclear proteins (*TNP1, TNP2*) and finally by protamines *PRM1, PRM2* and *PRM3* (van Roijen et al. 1998; Braun 2001). In contrast to sperm from other species, mature human sperm still contain a significant amount of histones, including testis-specific histone 2B (TH2B) (van Roijen et al. 1998). Expression and knockout studies in mouse underline the essential role of sperm chromatin genes and proteins in male fertility. These studies show that appropriate expression levels of *Prm1* and *Prm2* are required for male fertility and deletions of *Tnp1* and *Tnp2* cause subfertility, whereas mice with double knockouts for protamines and transitional nuclear proteins are completely infertile (O'Bryan and de Kretser 2006; Ravel et al. 2007). Mouse models also show that male fertility depends on the right balance between the sperm chromatin genes. For example, deletion of *Tnp1* triggers compensatory rise of the expression levels of *Prm2* and *Tnp2* and results in abnormal rod-shaped chromatin condensation, spermatozoa with blunted head tips and poor motility (Yu et al. 2000; O'Bryan and de Kretser 2006; Ravel et al. 2007). A heterozygous SNP in *PRM1* gene might be associated with male infertility also in humans (Iguchi et al. 2006).

iv) Capacitation, acrosome reaction and fertilization

The final and biologically most significant events in a sperm's life take place in the female genital tract and involve sperm capacitation, sperm-oocyte recognition, sperm-zona binding, acrosomal exocytosis (also known as acrosomal reaction) and culminate with fertilization. Capacitation hyperactivates sperm by giving it elevated motility to successfully penetrate egg's *zona pellucida* (ZP). It also destabilizes sperm's plasma membrane and prepares it for the acrosomal exocytosis. The latter happens when sperm contact the egg and the ZP proteins bind to the sperm plasma membrane (Mayorga et al. 2007; Varner and Johnson 2007; Zhao et al. 2007; Zhao et al. 2008). Acrosomal exocytosis facilitates the passage of the sperm through the zona and is absolutely necessary for fertilization (Mayorga et al. 2007; Varner and Johnson 2007). All these events are interrelated, regulated by complex biochemical pathways involving interactions of multiple proteins and protein complexes (De Blas et al. 2005; Mayorga et al. 2007; Varner and Johnson 2007). The best known of these is SNARE – a protein complex that controls membrane fusion and exocytosis, both in neuronal synapses and the sperm acrosome (De Blas et al. 2005; Kitamura et al. 2005). During recent years, the physiology and biochemistry of sperm capacitation and sperm-zona interactions have been extensively studied in mice, rats, humans and several domestic animals, including the horses (Varner et al. 2000; Bosard et al. 2005; De Blas et al. 2005; Neild et al. 2005; Obermann et al. 2005; Conner et al. 2007; Mayorga et al. 2007; Varner and Johnson 2007). These studies have revealed a number of previously unknown critical proteins such as complexins I and II that facilitate acrosome membrane fusions in the mouse

(Zhao et al. 2007; Zhao et al. 2008), or membrane-permeant protein *RAB3A* which is needed to trigger acrosome exocytosis in humans (Lopez et al. 2007). Contrary to this, very little is known about the genetic regulation of these processes. Discovery of a few candidate genes and causative mutations has been successful mainly in the mouse and rat because of the possibility to study changes in gene expression at different developmental stages, induce targeted mutations and generate knockout models or transgenic animals. For example, approximately 30 murine genotypes are known to be associated with impaired acrosomal exocytosis and 8 genotypes which cause errors in other aspects of sperm-oocyte interactions (MGI: http://www.informatics.jax.org/[1]). As expected, among the genes causing impaired acrosomal exocytosis (IAE) are those involved in cation channels and calcium oscillations (Fukami et al. 2001; Jin et al. 2005), tetraspanins that regulate membrane fusions (Tanigawa et al. 2008) and acrosome expressed proteins (Lee et al. 2008), but also neurotransmitter receptors because the biochemical events in acrosomal exocytosis closely resemble those taking place in neuronal synapses (Sato et al. 2000; Meizel and Son 2005). Precise molecular functions of these and several other candidate genes are yet to be discovered and mouse models suggest that the events from sperm capacitation to fertilization are controlled by synergistic interactions of many genes (Nayernia et al. 2003). In summary, genetic regulation of male developmental pathway from sex determination to fertilization in mammals is complex and involves interactions of many genes. Recent discoveries that spermatogenesis is regulated also by epigenetic factors and that aberrant DNA methylation in control regions of imprinted genes expressed in sperm might be associated with human oligozoosperma (Filipponi

and Feil 2009) indicate that the regulatory machinery governing male fertility is probably more complicated than currently appreciated.

Chromosomal distribution of male fertility genes in mammalian genomes

Mammalian male fertility is influenced by thousands of genes which are distributed throughout the genome. In order to target the likely candidate genes more efficiently, it is important to know whether this distribution is uniform between and within chromosomes or if certain genomic regions are more enriched for male fertility genes than others. Furthermore, the position of a gene in the genome might have important consequences for its function (Betran et al. 2004). With the advancement of gene mapping and sequencing technology, medium- to high-resolution gene maps and/or whole genome (WG) sequence data are available for more than 50 mammalian species including human, chimpanzee, model species such as mouse and rat, and a number of domestic animals, *viz.*, cattle, sheep, pig, horse, dog, cat, and rabbit (http://www.ncbi.nlm.nih.gov/[2], http://www.ensembl.org/index.html[3], http://genome.ucsc.edu/[4]). However, to date, identification and mapping of candidate genes for male fertility has been successful mainly in mouse and to a limited extent in humans. The available data from these two species indicate that the majority of male fertility genes are located on autosomes and the X chromosome, while a small proportion is strictly male specific and reside on the Y chromosome (Charlesworth 1991; Charlesworth and Charlesworth 2000; Charlesworth 2002; Charlesworth and Charlesworth 2005; Gvozdev et al. 2005; Graves et al. 2006). A comprehensive review on mammalian fertility genes

was published recently by Matzuk and Lamb (2008) and describes over 460 autosomal and X-linked murine genes of which 208 are specifically associated with male fertility. Additionally, the authors list over 200 human genes that are associated with different human male infertility phenotypes. The review provides information about the genes, mutations (if known), corresponding reproduction phenotypes, fertility status (e.g., subfertile, infertile) and chromosomal location. For example, there are about 40 genes responsible for spermatid differentiation. Of these *Adamts2* and *Ube2b* are located on mouse (*Mus musculus*, MMU) chromosome 11, *Ddx25* on MMU9 and *Six5* on MMU7. Human orthologs of these genes are located on human (*Homo sapiens*, HSA) autosomes HSA5, HSA11 and HSA19, respectively. Over 1000 genes are reported to be associated with spermatogenesis. To name a few, estrogen receptor 1 (*ESR1)* is located on MMU10 and HSA6, c-kit oncogene (*KIT)* on MMU5 and HSA4, and *kit ligand* (*KITLG)* on MMU5 and HSA12, respectively (Galan et al. 2005; Galan et al. 2006). Overall, the distribution of male fertility genes among autosomes is relatively uniform with slightly higher numbers of genes located on gene rich chromosomes such as MMU11 and HSA17 (Matzuk and Lamb 2008).

Several studies have underlined a special connection between the X chromosome and male fertility genes (Betran et al. 2004; Emerson et al. 2004; Wang 2004; Ellis and Affara 2006). First, several autosomal genes and gene families with testis-biased expression have been retrotransposed from X-linked genes (Emerson et al. 2004). For example, human *CSTF2T* is the retrotransposed autosomal gene originating from its X-linked progenitor *CSTF64*. *CSTF2T* shows testis-specific expression and has been

mapped to HSA10q22-23 – a region involved in reciprocal translocation in oligozoospermic males (Wang 2004). A mutation in an X-derived autosomal retrogene *mUtp14b* causes *juvenile spermatogonial depletion* (*jsd*) phenotype in mouse (Rohozinski and Bishop 2004). Second, X chromosome has a tendency to recruit genes involved in male reproduction from other genomic locations and is therefore disproportionately enriched with testis-specific genes but also with genes involved in brain function, signifying the role of brain in male reproductive success (Zechner et al. 2001; Emerson et al. 2004; Wang 2004; Ellis and Affara 2006). Alternatively, accumulation of testis and brain genes on the mammalian X chromosomes might be because the same protein complexes, thus the same genes, regulate interactions between the sperm and the egg and between neurons in the brain (De Blas et al. 2005; Kitamura et al. 2005). It is likely that due to hemizygous condition in males, the X chromosome genes are under unique evolutionary pressure and hence, the X chromosome serves as a preferred location for spermatogenesis genes (Torgerson and Singh 2006). Some of the X-linked spermatogenesis genes, such as the *Rhox* (Reproductive Homeobox genes on X chromosome) gene family, have been so far discovered only in mice. These genes are present on the X chromosome in clusters, predominantly expressed in Sertoli cells of adult testis and are involved in spermatogenesis (MacLean et al. 2005; Daggag et al. 2008). It has been common knowledge that X-linked spermatogenesis genes are expressed before sex chromosome meiotic silencing (Reinke 2004). Therefore, the recent discovery that 33 mouse X-linked multicopy gene families, representing approximately 273 spermatogenesis genes, are expressed predominantly in post-meiotic

cells (Mueller et al. 2008) is of outstanding importance and might fundamentally change our understanding about the function of reproduction related genes during male meiosis.

The mammalian Y chromosome is typically the smallest element in the karyotype, and forms only about 2% of an average mammalian genome. Surprisingly, despite the minute size and small proportion of euchromatin compared to heterochromatic regions, the Y chromosome is highly enriched with male fertility genes. However, to grasp a better understanding of how genes present on the Y chromosome acquired male fertility related functions, it is necessary to appreciate the evolution of this unique chromosome. A detailed knowledge about Y chromosome studies in human and mouse and the role of Y chromosome genes in male fertility in these two species will underline the importance of the Y chromosome in mammalian male fertility. The following sections will present a comprehensive overview about the evolution, structure and function of the mammalian Y chromosome.

MAMMALIAN Y CHROMOSOME AND ITS ROLE IN MALE FERTILITY

The origin and evolution of the Y chromosome

An hypothesis accounting for mammalian sex-chromosome evolution was first proposed by Susumo Ohno in 1967 suggesting that sex chromosomes in mammals evolved from a pair of ancestral autosomes about 300 million years ago (Ohno 1967). In course of evolution, one of the homologs in this ancestral chromosome pair acquired mutations that possibly had advantage in males. These mutations might have given rise

to a male sex determining locus, TDF (testis-determining factor, later known as Sex determining Region on Y, *SRY*) on one homolog which then became proto-Y chromosome, designating the other homolog as the proto-X chromosome. The proto Y started accumulating male advantage genes around the TDF, suppressing recombination with its homolog to keep a male-specific gene package together, and thus, creating a male-specific region on the Y chromosome (MSY). Suppression of recombination between X and Y initiated differentiation of these two chromosomes about 240-320 million years ago, shortly after the divergence of mammalian and avian lineages (Lahn and Page 1999). Thereafter the Y chromosome rapidly started to accumulate mutations and deletions which were not eliminated efficiently by selection pressure due to gradual loss of recombination with the X chromosome (Charlesworth 1991). As a consequence, in course of millions of years of evolution the male specific and hemizygous Y chromosome lost the majority of the ancestral set of genes and acquired an extensive amount of heterochromatic material. It retained only the genes that had selective advantage for critical male-specific functions and were possibly detrimental to females (Graves 1995, 2004, 2006; Delbridge and Graves 2007). The present mammalian Y chromosome is typically one of the smallest elements in the genome, has lost about 1250 of its original ~1300 genes and is a degraded version of the X chromosome. The outstanding evolutionary history of the Y chromosome thereby explains its unique specialization for genetic regulation of male reproduction.

Due to the presence of extremely high amounts of repeats, the Y chromosome was regarded, until as late as mid-1980s, as a genetic wasteland with the sole function of

male sex determination (Polani 1981; Sinclair et al. 1990; Quintana-Murci and Fellous 2001) and largely ignored by human, mouse and other mammalian genome mapping projects. However, it soon became evident that the Y chromosome has retained more genes than only *SRY* and that a few of them regulate spermatogenesis and male fertility (Chandley and Cooke 1994; Lahn and Page 1997, 2000; Saxena et al. 2000; Makrinou et al. 2001; Dorus et al. 2003; Giachini et al. 2005; Lardone et al. 2007a). These pioneering studies opened a new avenue for Y chromosome research in mammals and eventually led to sequencing of the Y chromosome in humans.

Y chromosome studies in mammals

Human (*Homo sapiens,* HSA)

The human Y chromosome (HSAY), is approximately 60 Mb in size and largely (~60%) heterochromatic. The 23 Mb of euchromatic portion covers approximately 8 Mb on the short arm (Yp) and 14.5 Mb on the long arm (Yq) of HSAY. The 40 Mb of heterochromatin comprises the bulk of the distal part of HSAYq and approximately 1 Mb at the centromere. Besides large numbers of heterochromatic repeats, another signature feature of HSAY is the presence of sequences with multiple copies also known as ampliconic sequences. The ampliconic region on HSAYq is comprised of eight massive palindromes and is one of the most peculiar structural features of the chromosome (Skaletsky et al. 2003).

Complete sequencing of the male-specific region of the human Y chromosome (MSY) revealed the presence of 156 transcriptional units which include 78 functional genes that collectively encode at least 27 distinct proteins. Several MSY genes are members of multicopy gene families of which 10 are expressed specifically in testis (Skaletsky et al. 2003). The HSAY gene list is still not complete because only recently eight more genes, all with open reading frames (ORFs), were discovered from a small euchromatic island within the pericentromeric heterochromatin (Kirsch et al. 2005). According to evolutionary origin and copy numbers the human Y chromosome genes are divided into four categories: X-degenerate, ampliconic, X-transposed and pseudoautosomal genes (Lahn and Page 1997; Skaletsky et al. 2003).

X-degenerate genes share ancestral homology with the X chromosome and the present-day HSAY has retained 16 genes with a gametologue on the X chromosome, for example *UTY-UTX* or *USP9Y-USP9X*. X-degenerate genes are usually single copy and are expressed in most of the body tissues including testis, *i.e.,* their expression is ubiquitous. They are usually involved in housekeeping activities but might have gained also other functions. The most remarkable example of X-degenerate genes is the male sex determination gene *SRY* which shares ancestral homology with X-linked *SOX3* but has gained entirely new and unique functions on the Y chromosome (Skaletsky et al. 2003; Sekido and Lovell-Badge 2008; Hiramatsu et al. 2009).

Ampliconic genes, as inferred from the name, are present in multiple copies. They can be Y-borne (present only on Y) or have moved to the Y chromosome from other genomic locations. It is believed that amplification of these sequences on the Y

chromosome has occurred mainly through gene conversion and duplications (Skaletsky et al. 2003). Demarcation between X-degenerate and ampliconic genes is not always distinct. Some X-degenerate genes such as *TSPY* (X-linked homolog is *TSPX*) or *RBMY* (X-linked homolog is *RBMX*) can also be present in multiple copies (Skaletsky et al. 2003; Delbridge et al. 2004). The majority of ampliconic genes, however, show testis-limited expression and might be involved in spermatogenesis and other male reproductive functions (Prosser et al. 1996; Lahn and Page 1997; Mitchell 2000; Toure et al. 2004a; Toure et al. 2004b; Delbridge and Graves 2007). It has been hypothesized that these genes have acquired multiple copies due to significant role in male fertility, so that a potentially harmful mutation in one copy will be compensated by multiple other normally functioning copies (Graves 2006). Such compensatory mechanism is necessary to grant normal function of genes on the hemizygous and non-recombining Y chromosome. A total of 9 ampliconic gene families with approximately 60 transcriptional units are present on HSAY.

X transposed genes, *TGIF2LY* and *PCDH11Y,* have moved to HSAY from HSAXq21 (Skaletsky et al. 2003) and are characteristic to human Y chromosome only. No X transposed genes have yet been found in other mammals including the closest relative of human, the chimpanzee (Hughes et al. 2005; Kuroki et al. 2006).

Pseudoautosomal genes are present in the pseudoautosomal region (PAR) which is a short region of homology between the mammalian X and Y chromosomes. The pseudoautosomal genes behave like autosomal genes by pairing and recombining during meiosis (Burgoyne 1982) and are thus inherited in autosomal rather than sex-linked

pattern (Mangs and Morris 2007). Human sex chromosomes are exceptional among mammals by having two pseudoautosomal regions - PAR1 and PAR2 (Goodfellow et al. 1986; Rouyer et al. 1986; Freije et al. 1992). PAR1 corresponds to other eutherian PARs while PAR2 is human specific. PAR1 is approximately 2.6 Mb, located at the terminal region of the short arms of the X and Y chromosomes (Goodfellow et al. 1986; Rouyer et al. 1986; Rappold 1993; Graves 1998), whereas PAR2 is a fairly short region spanning only 320 kb at the tips of the long arms of the sex chromosomes (Freije et al. 1992). PAR1 consists of at least 24 genes that have been characterized so far (Mangs and Morris 2007) while PAR2 contains only 4 genes, *viz., SPRY3, SYBL1, IL9R* and *CXYorf1* of which none has been associated with likely fertility functions (Kvaloy et al. 1994; Li and Hamer 1995).

Mouse (*Mus musculus,* MMU)

Besides human, mouse is the only other mammalian species with detailed information about the organization, gene content and gene expression profiles for the Y chromosome, MMUY (Mazeyrat et al. 1998; Affara and Mitchell 2000). Murine Y chromosome studies strongly confirm the findings in humans and show that MMUY contains transcriptionally active genes, a number of which are critically involved in male fertility and reproduction (McElreavey et al. 2002; Toure et al. 2004a; Toure et al. 2004b; Ellis et al. 2007; Grzmil et al. 2007). Currently, 53 genes and gene families are assigned to MMUY (http://www.ncbi.nlm.nih.gov/mapview/[5]). Like in humans, mouse Y chromosome genes are either X-degenerate, belong to multicopy gene families or are

pseudoautosomal. Notably, MMUY multicopy gene families are murine-specific, share no homology with human ampliconic genes but, interestingly, carry out similar male fertility related functions (Ellis et al. 2007). MMUY mutations have directly been associated with different male infertility phenotypes (Cooke and Saunders 2002). In contrast to human, mouse pseudoautosomal region (PAR) is only ~ 700 kb in size, harbors only two genes (*STS*, *MID1*) and shares no homology with human PAR1 (Perry et al. 2001).

These outstanding findings in human and mouse have encouraged Y chromosome studies also in other mammalian species and will be discussed in the following paragraphs.

Chimpanzee (*Pan troglodytes,* PTR)

Beside human, chimpanzee is the only species where the sequence information of Y chromosome is available. The sequencing of PTRY involves 12.7 Mb of MSY and 271 kb of PAR. Sequencing of the X-degenerate portion (~9.5 Mb) of the chimpanzee MSY has identified 19 protein-coding and 13 transcribed genes (Hughes et al. 2005; Kuroki et al. 2006). 59 pseudogenes were identified in the analyzed region of PTRY. Notably, all genes annotated as pseudogenes in humans are also pseudogenes in chimpanzee with one exception *CD24L4* which is a pseudogene in human is absent in chimpanzee (Kuroki et al. 2006). Despite the similarities, there are also pronounced differences between the two species - four PTRY genes *viz., CYorf15a*, *CYorf15b*, *TBL1Y*, *TMSB4Y* and *USP9Y*, have disruptions and mutations causing alteration or loss

of function (Hughes et al. 2005; Perry et al. 2007). In contrast to human where X-degenerate sequences are distributed along both arm of the Y chromosome and they are interrupted at several points by large blocks of ampliconic, heterochromatic and other sequences, the X-degenerate sequences in chimpanzee are present in a single, nearly contiguous block on the long arm of the PTRY. Moreover, there are also differences in the repeat elements. For example, PTRY harbors active L1 elements and endogenous retroviruses whereas HSAY is more enriched with Alu elements. Two novel endogenous retoviruses, CERV1 and CERV2 are present in 21 copies on PTRY but completely absent from human genome (Hughes et al. 2005). Despite these differences, overall, HSAY and PTRY have a sequence divergence of 1.78% whereas for rest of the genome the sequence divergence between human and chimpanzee is 1.23% (Kuroki et al. 2006).

Cattle (*Bos taurus,* BTA)

The first radiation hybrid (RH) map of the bovine Y chromosome (BTAY) was generated in 2002 (Liu et al. 2002) using a 7000 rad panel of 92 cattle x hamster hybrid clones. The map consisted of 49 microsatellites, 3 genes and 10 Sequence Tagged Sites (STSs). Microsatellite and STS markers were developed from a bovine Y-specific library, constructed by chromosome microdissection and microcloning. Two years later three more genes, *SRY*, *ANT3* (currently known as *SLC25A6*) and *CSF2RA* were mapped on BTAY by RH and FISH (Liu and de Leon 2004). The latter two genes and another gene, *STS*, were assigned to the bovine PAR (Moore et al. 2001). Recently the bovine pseudoautosomal boundary (PAB) was identified in the *SHROOM2-GPR143* interval.

The cattle PAR is longer than the human PAR but fewer genes have been indentified in cattle PAR so far compared to human PAR and the PAB extends more towards the centromeric region of the BTAY (Van Laere et al. 2008). Polymorphism analysis of 38 BTAY microsatellites showed that 24 have two or more alleles in cattle populations, whereas 7 microsatellites showed the presence of multiple copies in the same individual – a typical feature of Y-linked multicopy sequences (Liu et al. 2003).

Pig (*Sus scrofa,* SSC)

The only available porcine Y chromosome map (Quilter et al. 2002) comprises of 10 Y-specific genes. The gene order was determined by RH analysis and later confirmed by dual-color FISH. Both RH mapping and FISH confirmed that the genes on SSCY are clustered in two groups on the porcine MSY: a distal group with 6 genes (*AMELY, EIF2s3Y, ZFY, USP9Y, DBY* and *UTY*) and a proximal group with 4 genes (*TSPY, SMCY, UBE1Y* and *SRY*). As in humans, mice and cats, the order of *USP9Y-DBY-UTY* has been evolutionarily conserved also in pigs. Since the relative order of X-Y gametologues on porcine sex chromosomes is conserved, it was proposed that the porcine X-Y homologous genes closely resemble the ancestral eutherian sex chromosome gene order (Quilter et al. 2002).

Cat (*Felis catus,* FCA)

The basic RH map of cat Y chromosome was first published in 1999 (Murphy et al. 1999a). The map comprises of 8 genes and provides a preliminary comparison with

the existing human and mouse Y chromosome maps. Three loci, *USP9Y-DBY-UTY*, showed evolutionarily conserved physical order across all three species (Murphy et al. 1999b). As mentioned in the previous section, conserved linkage between these three genes was observed also in pigs (Quilter et al. 2002). It is noteworthy that the region containing *USP9Y-DBY-UTY* on FCAY corresponds to the AZFa region in humans and to Sxr^b interval in mouse where deletions are associated with spermatogenic arrest (Mazeyrat et al. 1998; Luddi et al. 2009). The most recent FCAY RH map contains 12 single copy X-degenerate genes (Murphy et al. 2006). Additionally, 6 multicopy Y-linked genes were identified using cDNA selection procedure where flow sorted Y chromosome DNA was hybridized to testis cDNA (Murphy et al. 2006). Expression profiles of all feline Y-linked genes have been studied on a panel of selected body tissues and full length cDNA sequences have been generated for 10 X-degenerate FCAY genes (Pearks Wilkerson et al. 2008). After human and mouse, FCAY is to date one of the best studied mammalian Y chromosomes.

Dog (*Canis familiaris,* CFA)

One of the first studies involving the dog Y chromosome described cloning and mapping of two pseudoautosomal genes, *CSF2RA* and *ANT3* (currently known as *SLC25A6*) on the dog sex chromosomes (Toder et al. 1997) and showed that canine PAR is larger than human PAR1 and extends further proximal on the short arm of the X chromosome. Later 10 male specific markers including the *SRY* gene were assigned to the canine Y chromosome using FISH and RH mapping approaches (Olivier et al. 1999;

Guyon et al. 2003b). Recent systematic gene discovery in the cat Y chromosome (Murphy et al. 2006) revealed that an X-degenerate gene *CUL4BY* and a Y-linked gene of autosomal origin *TETY2*, are both present in multiple copies also on the dog Y chromosome. Notably, these two genes have been found to be Y-linked so far only in carnivores (dog and cat) (Murphy et al. 2006). Additionally, 14 SNPs have been identified from CFAY sequences and used for population studies in dogs to generate the first dog Y chromosome phylogeny (Natanaelsson et al. 2006).

Rat (*Rattus norvegicus,* RNO)

Compared to the extensive Y chromosome research in mouse, the studies of the rat Y chromosome have been quite modest. Nevertheless, a few interesting differences have been discovered between the Y chromosomes of the two closely related species. As in most other mammals, the rat *Tspy* is a functional gene whereas mouse *Tspy* has become a nonfunctional pseudogene (Mazeyrat and Mitchell 1998). Studies show that rat *Tspy* is a low copy number gene with at least two copies on RNOYp – one copy being complete and functional, another truncated and probably nonfunctional (Dechend et al. 1998). Similarly to other species, rat *Tspy* demonstrates testis-specific expression (Mazeyrat and Mitchell 1998). Another RNOY gene that has attracted attention is *SRY*. While *SRY* in mouse and human is a single copy gene, in rat it is present in multiple copies. At least six full length copies of *Sry* have been identified on RNOY. These copies have a consereved coding region and conserved amino acid sequence (Turner et al. 2007). Expression analysis using RT-PCR showed that multiple copies of rat *Sry* are

expressed in adult testis and adrenal glands, though the possible function of the sex determining gene in adult testis is as yet poorly understood.

Rabbit (*Oryctolagus cuniculus,* OCU)

Y chromosome studies in rabbit have been limited mainly to the *SRY* gene. The 2,388 bp of rabbit *SRY*, including the open reading frame has been sequenced and two repetitive sequences in the 5′ region have been identified (Geraldes et al. 2005). Additionally, a 7 bp insertion polymorphism was recently discovered in the 3′ untranslated region and it was shown that the rabbit *SRY* gene is duplicated and evolves under concerted evolution (Geraldes and Ferrand 2006). *SRY* functional studies show that *SRY-SOX9* expression regulates gonadal morphogenesis in rabbit the same way as in human and other mammals but is different from mouse (Diaz-Hernandez et al. 2008).

Summary of Y chromosome studies in mammals

The data hitherto available on Y chromosomes of different mammals suggest that the chromosome carries functional genes but the gene content is not strongly conserved across species (Ehrmann et al. 1998; Liu et al. 2002; Ma et al. 1993; Mazeyrat et al. 1998; Mitchell et al. 1991; Murphy et al. 1999b; Quilter et al. 2002). There is a small core set of X-degenerate genes which is shared between mammals, while most of multicopy genes are restricted to one species or a group of related species. Thus, through species-specific loss and acquisition of genes, mammalian Y chromosomes have acquired unique repertoires of genes and it is not uncommon that Y-linked orthologs

exhibit differences in copy numbers and expression patterns (Delbridge et al. 1999; Graves 1995). For example, *TSPY*, is a testis-specific multicopy gene in human and primates, rat, cat, cattle, goat, sheep and horse (Arnemann et al. 1987; Dechend et al. 1998; Manz et al. 1993; Mazeyrat et al. 1998; Murphy et al. 2006; Raudsepp et al. 2004; Vogel et al. 1997a; Vogel et al. 1997b) but a nonfunctional single copy gene in mouse (Mazeyrat et al. 1998; Schubert et al. 2000a; Schubert et al. 2000b). These unique properties of the Y chromosome set limitations to the use of comparative genomics in Y chromosome research and necessitate species-specific approaches. Therefore, in order to develop a proper understanding of the expression and function of Y-linked genes, it is important to identify and analyze these genes separately within each species.

Mammalian Y chromosomes, though variable in organization and gene content among species, share some important common features of which the most distinctive is the presence of multicopy sequences. It is proposed that due to the lack of recombination over most of its length, amplification of genes may be an important compensatory mechanism that protects functionally essential sequences from being lost through mutational degradation (Skaletsky et al. 2003). Since most of multicopy genes typically show testis-specific expression it is likely that they have roles in regulating male fertility. Therefore, the identification and functional analysis of Y-linked multicopy sequences might be a key to discover important male fertility genes in mammals.

Male fertility and mammalian Y chromosome genes

In 1931, eminent evolutionary biologist and geneticist Sir Ronald Fisher suggested that genes involved in benefit to the male (including spermatogenesis genes) would accumulate on the Y chromosome (Affara and Mitchell 2000). The validity of these ideas was proved approximately 45 years later by the discoveries that human and mouse Y chromosome microdeletions and mutations are associated with diverse spectrum of defective spermatogenic phenotypes (Cooke and Saunders 2002; Mazeyrat et al. 1998; Tiepolo and Zuffardi 1976). Different portions of human and mouse MSY show the presence of intervals containing one or more genes that control male germ cell differentiation and hence, are capable of causing infertility. These intervals have been mapped, cloned and examined in detail for functional genes. For example, deletion of the distal portion of HSAYq11, a region known as azoospermia factors (AZFa, AZFb, AZFc) containing clusters of spermatogenesis genes causing azoospermia in humans (Tiepolo and Zuffardi 1976; Vogt et al. 1996). It has been shown that the deleted region of *AZFc* in infertile men, always includes *DAZ* (deleted in azoospermia) genes and the frequencies of Y chromosome microdeletions and chromosomal abnormalities are associated with azoospermic and oligozoospermic conditions in males (Vicdan et al. 2004). Additionally, it has been shown that distinct regions in HSAYq11 overlapping with the genomic AZFb and AZFc intervals are probably involved in the pre-meiotic X and Y chromosome pairing process (Vogt et al. 2008). Murine Y chromosome interval ΔSxr^b contains *Spy* spermatogenetic factor genes, the expression of which is important for proper development of germ cells. Deletion of this portion of ΔSxr^b impairs fertility

in mouse. Studies indicate that AZFa in human and *Spy* in mouse are both critical for male fertility and may be encoded by homologous genes (Mazeyrat et al. 1998).

Although Y chromosome microdeletions account for the most important genetic causes of spermatogenic failure in humans, a significant proportion of patients with impaired spermatogenesis are not known to have Y deletions. Studies in human and mouse show that male infertility can be caused also by other specific changes in a number of Y-linked genes. A few of them will be discussed briefly in the following sections.

RBMY is a member of RBM (RNA-binding motif) family and encodes a protein that localizes to the nucleus of all spermatogenic cell types (Elliott et al. 1996; Elliott et al. 1997; Elliott et al. 1998). *RBMY* is required for normal male fertility, but because *RBMY* is a multicopy gene, it is not easy to understand the effect of *RBMY* deletions on fertility (Kostiner et al. 1998). However, there is an indirect evidence of *RBMY* in relation to male fertility, based primarily on the correlation between the AZFb interval and expression of the gene encoding the RBMY protein in germ cells. Human *RBMY* is located in the AZFb interval and deletion of this region abolishes RBMY protein production in germ cells of affected males (Affara 2001). Since human RBMY protein recognizes RNA through a novel mode of protein interaction which could be important for the initiation of spermatogenesis (Skrisovska et al. 2007), the lack of RBMY protein might seriously affect the production of germ cells. Mouse studies, on the other hand, show that severe *Rbmy* deficiency causes increase in the number of abnormal sperm (Szot et al. 2003). Twelve potential target mRNAs that are bound to the RBMY protein

have been identified in mouse testis. Murine RBMY protein can bind also to its own transcript and may, thus, affect alternate splicing and regulate its own expression. These are new findings and it is necessary to further explore the role of *RBMY* in spermatogenesis (Zeng et al. 2008).

DAZ (**D**eleted in **AZ**oospermia) gene cluster was transposed to the Y chromosome during primate evolution but has remained at its original autosomal location in other mammalian species (Saxena et al. 1996; Saxena et al. 2000; Skaletsky et al. 2003). The gene cluster is named "deleted in azoospermia" (DAZ) because deletion in this region completely removes all *DAZ* genes in an individual and causes azoospermia. *DAZ* has a diverse spectrum of transcripts reflecting the transcriptional activity of all members of the gene family (Yen et al. 1997). Functional homologs of human DAZ genes have been found also in mouse Y chromosome and are called *Dazl* (*DAZ*-Like). Mice with no functional *Dazl* genes are sterile, whereas heterozygous mice with one intact copy of *Dazl* are subfertile due to the reduced number of germ cells. This indicates that the number of *DAZ* or *Dazl* copies might be important for germ cell production (Kostiner et al. 1998). This is supported by human studies showing that three different patterns of *DAZ* gene deletions are associated with severe oligozoospermia and infertility (A et al. 2006). In patients with spermatogenic arrest *DAZ* gene transcripts are localized in primary spermatocytes and *DAZ* gene activity seems to correspond to the proliferative activity of stem cells of germinal epithelium (Szczerba et al. 2006).

TSPY (testis-specific protein, Y encoded) genes are part of DYZ5 - a clustered array of repeats, located on the proximal short arm of the human Y chromosome.

Evolutionary conservation of *TSPY* sequence and expression pattern among primates suggests that *TSPY* might play an important role in testicular function and male fertility (Manz et al. 1993). Similarly to primates, *TSPY* genes are arranged in clusters on the Y chromosome of several other mammals and show sequence conservation across species (Mazeyrat and Mitchell 1998; Skaletsky et al. 2003). It is proposed that *TSPY* is involved in early stages of spermatogenesis to help the proliferation of spermatocytes (Schnieders et al. 1996; Vogel and Schmidtke 1998). Extensive amplification of *TSPY* on the Y chromosome of most studied mammals indicates that there might be a requirement for a critical dose of *TSPY* during spermatogenesis (Delbridge et al. 2004). Notably, infertile men have more *TSPY* gene copies compared to normal fertile men insisting that copy number difference might be the main factor that determines the effect of *TSPY* on spermatogenesis and male fertility in humans (Vodicka et al. 2007).

UBE1Y is a Y-linked gene in mouse and several other mammals (Mitchell et al. 1991; Quilter et al. 2002; Murphy et al. 2006) but has been lost from the Y chromosomes in humans, chimpanzees and old world monkeys (Mitchell et al. 1998). The fact that *UBE1Y* encodes for ubiquitin activating enzyme E1 which promotes cell proliferation and mitosis (Odorisio et al. 1996) makes it a strong candidate gene for spermatogenesis (Levy et al. 2000). This is supported by mouse studies showing that *Ube1y* is expressed exclusively in spermatogonial cells and is part of a gene cluster in the ΔSxr^b deletion interval known to be critical for the proliferation of mitotic germ cells during spermatogenesis (Levy et al. 2000).

USP9Y (Ubiquitin-specific protease 9 Y) is located in the AZFa region of human Y chromosome and possesses ubquitin protease activity (deubiquitination) by which it might stabilize specific target proteins that are important for male germ cell development (Lee et al. 2003). Data about the role of *USP9Y* in male fertility are controversial. There are studies showing that a point mutation in *USP9Y* causes partial arrest of spermatogenesis at the post meiotic spermatocyte stage resulting in severe spermatogenic failure and infertility (Sun et al. 1999). However, recently a complete deletion of *USP9Y* was found in a normospermic man, his brother and father without any deleterious effect on fertility (Luddi et al. 2009). This indicates that *USP9Y*, previously considered a candidate gene for male infertility and azoospermia, does not play a key role in male reproduction. It is likely that *USP9Y* acts rather as a fine tuner to increase efficiency of spermatogenesis than being a direct provider of essential functions (Krausz et al. 2006).

SRY, the sex determining region on Y, is indisputably the "signature" gene of the Y chromosome and has been studied, compared to other Y-linked genes, in more detail. Besides human and mouse, the coding sequence of this small, one exon gene is characterized in bovids (Cheng et al. 2001), horse (Hasegawa et al. 1999), rat (Turner et al. 2007), pig (Parma et al. 1999), sheep (Payen et al. 1996), goat (Pannetier et al. 2006) and rabbit (Geraldes and Ferrand 2006; Geraldes et al. 2005). *SRY* conserved regulatory regions have been identified in 10 different mammals (Margarit et al. 1998) and studied in detail in pigs (Boyer et al. 2006; Pilon et al. 2003). Information about SRY expression is available for human (Skaletsky et al. 2003), mouse (Mazeyrat et al. 1998), rat

(Mazeyrat and Mitchell 1998) horse (Hasegawa et al. 1999), pigs (Boyer et al. 2006; Pilon et al. 2003) and carnivores (Murphy et al. 2006). These studies indicate that *SRY*, despite of being the "master" sex determination factor in all eutherians, exhibits variation in coding and regulatory sequences, as well as in expression profile across species. Furthermore, *SRY* shows variation also in copy numbers being a single-copy gene in human and mouse and having multiple copies in cat, rat and rabbit (Nagamine 1994; Bullejos et al. 1999; Skaletsky et al. 2003; Geraldes et al. 2005; Turner et al. 2007; Pearks Wilkerson et al. 2008). The most intriguing feature of *SRY*, however, is that in most of the species studied it is highly expressed in adult testis indicating that male sex determination is not the only function of *SRY* and it might be involved in processes taking place also in adult testis.

In addition to Y chromosome microdeletions and gene mutations, some infertile phenotypes can be caused by transcriptional alterations of AZF genes. For example, infertile males with complete Sertoli cell only syndrome (SCOS) show no expression of *RBMY*, *DAZ* and *TSPY* whereas *DDX3Y* (alias *DBY*) is expressed at very low levels (Lardone et al. 2007b).

In summary, studies in human, mouse and other mammals provide convincing evidence that Y chromosome sequences are critically involved at various stages of spermatogenesis and play essential roles in regulating male fertility. The critical role of the Y chromosome in male fertility is further underscored by the facts that about 10-15% of idiopathic male infertility in humans is caused by Y chromosome mutations and that

the currently available genetic diagnostic tests for human male infertility are exclusively based on Y chromosome markers (Peterlin et al. 2004).

MALE FERTILITY IN HORSES

Stallion reproduction and fertility is of vital significance to the horse industry. There are over 9.2 million horses in the United States and it is estimated that almost one-fourth of horses are used for breeding purposes. Stud fees for stallions can range from a few thousand to half a million dollars (www.horsecouncil.org[6]). Thus, fixed expenses, lost potential profits and reduced revenue from impaired fertility of stallions has an enormous economic impact on the horse industry. Because of the structure of horse breeding where one stallion covers many mares, economic losses due to reduced fertility of stallions can be even more dramatic. In contrast to many other livestock species where animals are selected for fertility, stallions are selected mainly on the basis of their ancestry, athletic performance and overall appearance, but not for their reproductive potential (Colenbrander et al. 2003). Furthermore, not all stallions are subjected to a detailed breeding soundness examination prior to standing at stud (Colenbrander et al. 2003) and fertility problems become apparent only later when desired foaling rate is not achieved. Due to this, poor fertility of breeding stallions is a serious problem worldwide. For example, it has been estimated that 36%-43% of prospective breeding stallions fail breeding soundness evaluation (Blanchard and Varner 1997; Woods et al. 2000). These examinations traditionally include analysis of semen parameters such as sperm count, sperm motility and sperm morphology (Graham 1996; Madill 2002; Colenbrander et al.

2003; Love et al. 2003), examination of external genitalia and a general physical evaluation. Additionally, the sperm chromatin structure assay and the acrosomal responsiveness assay can be carried out to evaluate sperm functionality in more detail (Love and Kenney 1998; Varner et al. 2000). More indirect measures, like per-cycle pregnancy rate are also used to estimate stallion fertility (Colenbrander et al. 2003; Love et al. 2003) but these factors are influenced by management practice of both mares and stallions (Rousset et al. 1987; Love et al. 2003) due to which precise definition of fertility in stallions remains obscure.

Of the two main factors governing stallion fertility and reproduction, environment has been extensively studied and optimized (Roser 2001; Madill 2002; Merkies and Buhr 2004) while very little is known about the role of genetics. As yet, only a handful of reproduction and fertility related genes have been analyzed in horses (Ing et al. 2004; Leeb et al. 2005; Hamann et al. 2007) and even fewer have been studied for their function, interactions and mutations that could lead to aberrant conditions and impaired fertility. The remarkable progress in equine genome mapping (Chowdhary et al. 2003; Penedo et al. 2005; Swinburne et al. 2006; Raudsepp et al. 2008a) which ultimately led to the sequencing of the whole genome (Wade et al. 2007) provides a unique opportunity to initiate organized genome-wide search for reproduction related genes in horses. However, because the equine genome is sequenced from a female individual, the whole genome sequence will not furnish any information on the genes and transcripts residing on the Y chromosome, several of which are critical for male reproductive development and fertility. Hence, obtaining information on the structure,

gene content and functional profile of the Y chromosome will be essential to identify the Y-linked components of stallion fertility.

PRESENT STATUS OF THE HORSE Y CHROMOSOME RESEARCH

The overall information on the structure and organization of the horse Y chromosome (ECAY) is limited – a situation common to other livestock and companion animal species. Cytogenetic studies indicate that the sub-metacentric equine Y chromosome (ECAY) is one of the smallest elements in the genome (Power 1988), consists of a large heterochromatic region which covers the entire short and majority of the long arm, while the tiny euchromatic region is located at the distal end of ECAYq. The chromosome was first microdissected in 1999 (Raudsepp and Chowdhary 1999) to make a chromosome-specific paint that was used for comparative mapping on metaphase spreads of different equids. The first formal map for the Y chromosome was generated by assigning *STS-Y*, *SRY* and *ZFY* to the Y chromosome by somatic cell hybrid analysis (Shiue et al. 2000). Later *SRY* and *ZFY* were localized by FISH to the long arm of the chromosome (Hirota et al. 2001). The first detailed physical map of horse Y chromosome containing 7 contigs was constructed in 2004 (Raudsepp et al. 2004b). The map comprises of 73 overlapping Bacterial Artificial Chromosome (BAC) clones and shows the linear order of 100 Sequence Tagged Sites (STSs) and 9 genes. The BACs spanning the minimum tiling path (MTP) of each of the contig indicate a cumulative coverage of ~4 Mb which corresponds to almost 20-25% of the ~15 Mb euchromatic region of ECAY (Raudsepp et al. 2004b). Notably, contig II in this map is composed of

genes, markers and BAC clones that are all present in multiple copies - a hallmark of mammalian Y chromosomes.

Recently a detailed map was constructed for the horse pseudoautosomal region (PAR) and physically assigned to the terminal end of the long arm of the Y chromosome (Raudsepp and Chowdhary 2008). The map has 129 physically ordered markers (110 STSs and 19 genes) contained in 71 BAC clones that are arranged into 2 contigs spanning the region. Horse, cattle (Van Laere et al. 2008) and dog (Young et al. 2008) are to date the only domestic animals with precisely defined pseudoautosomal boundary (PAB). Interestingly, a ~200 kb region was discovered in the middle of the equine PAR that is also present in the male specific region of the Y (MSY). Duplication of this kind is a novel observation in mammals and it is not clear whether this can affect sex chromosome pairing and segregation in male meiosis. Overall, following human/chimpanzee and mouse, horse is the only species where PAR is described in such details. The initial map of MSY and high-resolution map of PAR serve as an important foundation for further expansion of Y chromosome studies in horses.

RATIONALE OF THE PRESENT STUDY

The significant economic impact of stallion reproductive performance to the equine industry and our limited knowledge about the genetic factors governing male fertility, underline the urgent need to initiate systematic studies of male fertility genes in horses. Given the outstanding role of Y chromosome in human and mouse spermatogenesis, the specific focus of this research is the horse Y chromosome. The

rationale is that if we obtain thorough knowledge about the organization and functional profile of the Y chromosome, this will help us to identify key-role Y-linked male fertility genes in horses. The central focus of the work is, thus, isolation, identification and expression analysis of Y chromosome genes and ESTs (Expressed Sequence Tags). This information will be used to improve the existing ECAY map and develop a comprehensive, high-resolution map for the horse Y chromosome. Analysis of differential expression of Y-linked genes between normal stallions and stallions with fertility problems will help identify the key-role male fertility genes in horses.

The **specific goals** of the project are 1) to develop a comprehensive physical map and gene catalogue of the horse Y chromosome and 2) identify Y-linked candidate genes responsible for stallion fertility. To achieve these goals, first new Y chromosome genes/ESTs will be isolated from horse testis. Human and mouse studies have shown that 99% of Y-linked genes are expressed in testis (Skaletsky et al. 2003). Therefore, testis should be the most informative tissue for Y chromosome gene discovery also in horse. A direct cDNA selection procedure (Lovett et al. 1991; Del Mastro and Lovett 1997) will be used for new gene discovery from horse testis by hybridizing testis cDNA with Y-specific BAC clones and with flow sorted ECAY. This will enable the specific retrieval of only Y-linked testis transcripts. The Y chromosome BAC clones will also be used for chromosome walking and isolation of new clones to fill the gaps on the current map and to expand the map over the entire MSY. Expression patterns and copy numbers of the newly identified genes/ESTs will be analyzed to identify candidate genes for stallion fertility.

The **long term goal** of this project is to obtain complete sequence information for the horse Y chromosome. Generation of a comprehensive ECAY map in this project will serve as a good foundation for this goal. Detailed analysis of ECAY gene content will serve also as a basis for studying the evolution of ECAY compared to other mammalian Y chromosomes. Functional analysis of candidate genes for stallion fertility will help to understand the genes and gene networks regulating male fertility and develop diagnostic tests for early discovery of potentially infertile phenotypes in horses.

CHAPTER II

DISCOVERY OF Y-SPECIFIC GENES FROM HORSE TESTIS AND GENERATION OF A DETAILED MAP FOR THE HORSE Y CHROMOSOME

INTRODUCTION

In order to understand the role of the Y chromosome in the biology of a species it is necessary to acquire detailed information about the Y chromosome organization and most importantly, identify all Y-linked genes and study their structure, expression profiles and possible functions.

The first systematic search for genes from the non-recombining region of the Y chromosome (NRY), later known also as the male-specific region of Y (MSY, (Skaletsky et al. 2003), was carried out in humans (Lahn and Page 1997). This study used testis, a single complex organ, as the source of Y-linked transcripts and identified 12 novel genes or gene families. According to expression profiles, evolutionary history and possible function, these human Y-linked genes were classified into two categories: i) genes that were expressed in many tissues including testis, had an ancestral homolog on the X chromosome and carried out house-keeping functions, and ii) genes that were expressed exclusively in testes, belonged to Y-specific gene families and were possibly involved in male fertility-related functions. The authors concluded that in contrast to other chromosomes there is a clear functional coherence of the gene content of the Y chromosome.

This first gene discovery from MSY was carried out using the direct cDNA selection procedure (Lovett et al. 1991) from human testis where testis cDNA served as a driver and a flow-sorted human Y chromosome cosmid library as a selector. Later, the same procedure allowed isolation of 12 Y-specific genes in the domestic cat (*Felis catus*, FCA, (Murphy et al. 2006; Pearks Wilkerson et al. 2008). Notably, six FCAY genes were novel, three were derived from putative autosomal progenitors and another three were X-degenerate.

In horses the very first basic gene map of ECAY was built using a somatic cell hybrid (SCH) panel and showed synteny between *STS-Y*, *SRY* and *ZFY* (Shiue et al. 2000). A year later *SRY* and *ZFY* were assigned to ECAY by fluorescence *in situ* hybridization (Hirota et al. 2001). Currently partial genomic and/or cDNA sequences are available for nine ECAY genes, *viz.*, *AMELY, TSPY, DDX3Y (alias DBY), USP9Y, UTY, TBL1Y, KDM5D (alias JARID1D, SMCY), SRY* and *ZFY* (Agulnik et al. 1997; Raudsepp et al. 2004b), though no systematic efforts have been made to generate a comprehensive gene catalogue for the horse Y chromosome. In this study the cDNA selection procedure which has been successfully applied for Y chromosome gene discovery in humans and cats, will be adopted, and used it to identify Y-specific ESTs from horse testis. Flow-sorted equine Y chromosome and pools of BAC clones from the ECAY contig map (Raudsepp et al. 2004b) will serve as two different types of selectors to capture horse testis cDNA sequences. Y-specific ESTs will be assigned to the existing ECAY contig map by STS content analysis. The cDNA sequences will also be used to identify new BAC clones for the expansion of the contig map.

OBJECTIVES

1. Isolate and identify Y-specific genes/ESTs from horse testis by cDNA selection procedure.
2. Isolate BAC clones containing the Y-specific genes/ESTs and map the BACs to the ECAY contig map by STS content analysis.
3. Use STS content analysis and chromosome walking to expand the existing ECAY contig map and bridge gaps where possible.

As a result, Y-specific expressed sequences and genes will be identified and used to construct a comprehensive, physically ordered gene map for the horse Y chromosome. The map will serve as a template for complete sequencing of equine MSY. ECAY-specific transcripts will be analyzed for expression profiles to discover potential candidate genes for stallion fertility. Details of the experimental design and the outcomes of the experiments are presented in the following sections.

MATERIALS AND METHODS

Extraction of messenger RNA (mRNA) from horse testis

Messenger RNA (mRNA) was isolated from normal adult horse testis using Fast Track 2.0 mRNA isolation kit (Invitrogen) following the manufacturer's instruction with minor modifications. Briefly, about 1 g of testis tissue was homogenized using lysis buffer and incubated for 1 hour at 45^{o}C to facilitate complete lysis of cells. Next, 950 µl of 5M NaCl was added to the solution and DNA was sheared using a 22 gauge needle

and a 20 ml syringe. After adding Oligo dT cellulose the solution was incubated for 1 hour to ensure the binding of oligo dT cellulose to poly A tail of mRNA. The solution was then centrifuged at 3000 x g for 5 min at room temperature to pellet the cellulose. Supernatant was discarded and the cellulose was resuspended in 20 ml binding buffer. The centrifugation step was repeated twice and the final pellet was resuspended in low salt buffer to help remove the unbound materials from the cellulose. This washing step was repeated for three times and finally the cellulose was resuspended in 800 µl of low salt buffer and transferred into spin columns supplied with the kit. The columns were centrifuged at 5000 x g for 10 sec at room temperature to bind the cellulose with the spin column membrane. The flow through was discarded, 500 µl of low salt buffer was added to the column and centrifuged at 5000 x g for 10 sec at room temperature. The process was repeated for three more times to wash the cellulose. The column was placed into a fresh tube and 200 µl of pre-warmed (65^{o}C) elution buffer was used to elute the RNA from the column. Elution was repeated twice. The eluted RNA solution was mixed with 60 µl of sodium acetate and 1ml of 100% ethyl alcohol and stored overnight at -80^{o}C. The solution was then thawed and centrifuged for 30 min at 4^{o}C at maximum speed (13000 x g). The supernatant was removed and the pellet was centrifuged again to remove all remaining droplets. The final mRNA pellet was dissolved in 30 µl of elution buffer. DNase treatment was carried out using Turbo DNase Kit (Ambion) following manufacturer's instructions. Quantity and quality of the isolated mRNA was evaluated using NanoDrop Spectrophotometer and Agilent Bioanalyzer (Agilent, CA) with the RNA 6000 Nano chip kit (Agilent, CA).

cDNA synthesis

Testis cDNA was synthesized using a combination of random primers and oligo dT primers with Superscript II reverse transcriptase provided by Superscript Choice System for cDNA Synthesis (Invitrogen) following manufacturer's instructions. Briefly, 4 µg of testis mRNA was used for cDNA synthesis with random hexamer and oligo dT primers. The first and second strands of cDNA were synthesized in two steps followed by phenol chloroform extraction and cDNA precipitation with ammonium acetate and absolute alcohol. The cDNA was ligated with adapter primers, PCR amplified and designated as primary cDNA.

cDNA selection

Isolation of Y-specific ESTs from horse testis was carried out by direct cDNA selection following the protocol of Del Mastro and Lovett (1997). The PCR amplified primary cDNA was first pre-annealed with horse Cot-1 DNA (Cot1:cDNA ratio ranged from 2:1 to 15:1) for 4 hours to block repetitive sequences. In some experiments blocking DNA contained also cDNA sequences from *UBE1Y* - the most abundant ECAY transcript in testis - to allow better hybridization of unique and less abundant sequences (see discussion for explanations). Horse Y chromosome specific probes were generated from two sources: i) flow sorted and GenomiPhi (Amersham Biosciences) amplified ECAY (7000 copies provided by Cambridge Resource Centre for Comparative Genomics) and ii) 183 BAC clones from the ECAY contig map (Raudsepp et al. 2004b; Raudsepp et al. 2008b). The BAC clones were divided into 6 pools with ~30 clones in

each. Flow sorted ECAY and BAC pools were labeled with biotin by nick translation (see FISH below) and hybridized individually with pre-annealed testis cDNA for 50 hours. Testis cDNA and biotinylated Y chromosome hybrids were captured with streptavidin coated paramagnetic beads (Dynabeads® M-280 Streptavidin, Invitrogen) and Y-specific testis cDNA was eluted. This primary selected cDNA was PCR amplified using adapter primers and used for a second round of hybridization to obtain secondary selected cDNA.

cDNA cloning, sequencing and sequence analysis

The secondary selected cDNA was PCR-amplified and cloned *en masse* into TOPO-TA cloning vector (Invitrogen) following manufacturer's recommendations. The vector was transformed into One Shot Top10 *E. coli* DH5α chemically competent cells (Invitrogen), plated on LB agar plates containing ampicillin (50 µg/ml) and incubated overnight at 37ºC. Plasmid clones were picked and grown overnight at 37ºC in 2 ml, 96-well culture plates containing LB media and ampicillin (50 µg/ml). Plasmid DNA was isolated with an alkaline lysis based kit (REAL-prep96; Qiagen) following manufacturer's instructions. A random set of clone DNA was digested with EcoRI (Invitrogen) to confirm the presence of different inserts. All cDNA clones were sequenced using BigDye (Applied Biosystems) terminator chemistry, universal primers (T7 and T3) and ~300-500 ng of plasmid DNA as template. The sequencing reactions were resolved on an ABI-3730 capillary sequencer (Applied Biosystems). After the sequences were quality trimmed and checked for vector contamination, they were

assembled into contigs using Sequencher V 4.7 software (GeneCodes Co). The contigs were checked for the presence of repetitive elements with RepeatMasker (http://www.repeatmasker.org[8]) and analyzed using Discontiguous MegaBLAST (http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi[9]) to identify putative orthologs from human, mouse and other mammalian genomes. The cutoff threshold was 1e-8. The sequences were also aligned with horse whole genome (WG) sequence assembly EquCab2 (http://www.ensembl.org/Equus_caballus/index.html[10]). Because a female horse has been sequenced, no alignment of WG sequences with the partial cDNA sequences indicated the likelihood that the selected cDNA sequences were Y-specific.

cDNA analysis by PCR

After analysis with Discontiguous MegaBLAST the sequences with similarity to mammalian X- or Y-linked orthologs, sequences with weak similarity to autosomal genes and sequences with no significant similarity to any mammalian sequences were considered for further analysis. Where possible, the partial cDNA sequences were aligned with human and mouse Y chromosome genes (http://www.ensembl.org/index.html[3]) to identify likely exon/intron boundaries. Intron-spanning or exonic primers (Table 1) were designed with Primer 3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi[11]).

Table 1: Detailed information about ECAY genes and ESTs.

Gene symbol	Gene name	Primers 5′ to 3′	Annealing temperature, °C	Genomic product size, bp	GenBank accession No
**AMELY*	Amelogenin, Y-linked	F: CCAACCCAACACCACCAGCCAA ACCTCCCT R: AGCATAGGGGGCAAGGGCTGCA AGGGGAAT	65	160	AB032194
CUL4BY	Cullin 4 B Y	F: TGTGGGGTTCGTGTGAAATA R: CAAGGATCGCTGGGTCTTAC	58	172	EU687546
CYorf15	Chromosome Y open reading frame 15	F:CAACCATGCATTGAAAGAGG R:TGCACTCCGATTCTTGTTGA	58	152	EU687545
DDX3Y (DBY)	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked	F:CTCGAGATCCAAAACTGCTG R:TGATAAAAACAGTTCAGGGTGGA	58	68	EU687547
EIF1AY	Translation initiation factor 1A Y	F:GATCGTGGCCTTCTGACATT R:TTATTTTTGGGCATGGTGGT	58	187	ET052957
**EIF2s3Y*	Eukaryotic translation initiation factor 2, subunit 3 gamma, Y- linked	F:GAGCCATCTGTGTGATCGTC R:TATTCCTGGCCCTAAGCACA	58	223	
EIF3CY (EIF3s8Y)	Eukaryotic translation initiation factor 3, subunit C on Y	F:CCCAAGCAGGGTACCTATGG R:GGACAGAAGTGACGCAATCA	58	134 (M), 230 (M/F)	EU687548
ETSTY1	Equus Testis-specific transcript Y1	F:GACGGACGACCTTGTGTTTT R:CTAGTGGCGAGTCCTTTTGG	58	234	EU687549
ETSTY2	Equus Testis-specific transcript Y2	F:ATCATCGTGGAAAGCCTCAC R:AGTGCTGAAGAGGCTGTGGT	58	223	EU687550
ETSTY3	Equus Testis-specific transcript Y3	F:TTACATTTGTTGCGCCATGT R:GCCCAAAGAAGTAACCGACA	58	134	EU687551
ETSTY4	Equus Testis-specific transcript Y4	F:TAAGGCTTCCCTCCTCCAAT R:CCAGTGACCCGACATACTGA	58	175	EU687552
ETSTY5	Equus Testis-specific transcript Y5	F:CAAAACCAAGAGGAGGACCA R:CTCCAGAGGCAGGTACTTCG	58	210	EU687553
ETSTY6	Equus Testis-specific transcript Y6	F:ACATGGCGCAACAAATGTAA R:TAGCTGTTTGCTGCAGTGCT	58	245	EU687554

Table 1 continued

Gene symbol	Gene name	Primers 5′ to 3′	Annealing temperature, °C	Genomic product size, bp	GenBank accession no
ETY1	Equus transcript Y1	F:TCCAGAGCAACAACAGCAAC R:CATCAGTCTGCCCAAACCTT	58	127	EU687555
ETY2	Equus transcript Y2	F:TAAGGCTTCCCTCCTCCAAT R:CCAGTGACCCGACATACTGA	58	850	EU687556
ETY3	Equus transcript Y3	F:TTTTGGCTTGTGTCTTTCTCTG R:ATAGGGCCAGACTTTCACAGC	58	150	EU687557
ETY4	Equus transcript Y4	F:TGGGGATATTGGCTTAGCTG R:CTGGGAGCACGTCTGTATCA	58	180	EU687558
**KAL1Y*	Kallmann Syndrome 1 on Y	F:AGGCACAGTCTTAGGGCAAA R: TTTTGGCATTCCCTTCTCTG	58	231	
KDM5D (SMCY)	Jumonji, AT rich interactive domain 1D	F: AACAGCGAGCCAATGTTTTT R: GCAAAATTCTGGGAAATCCA	58	400	EU687564
**MAP3K 7IP3Y*	Mitogen-activated protein kinase kinase kinase 7 interacting protein 3 on Y	F: GTGGAATCCCTATTGCTAAAGTTAC R: CCAGAGAGCTGTGACCAAG	58	138	
MT-ND1Y	Mitochondrially encoded NADH dehydrogenase 1 on Y	F:CCCTCCGCTTTCCTAGACC R:CAACGATGGCTTGAAAGGAT	58	100	EU687559
NLGN4Y	Neuroligin 4 isoform Y	F:GGGGATCCATCTTTGTGTTG R:GTCACACAGCAGGCTCTGAC	58	156	EU687560
RBMY	RNA-binding motif Y	F:TTCGGCCTTCTCTTTCACAT R:ACTCAAGCAGCCGAAATGAT	58	180	EU687561
RFX5Y	Regulatory factor X 5 on Y	F:ACCCTTAGGGGGAAAAATCC R:TTTCGTCCCTCAAGTTCCTG	58	201	EU687562
RPS3AY	Ribosomal Protein S3A	F:CCGGAAGAAGATGATGGAAA R:CAAACTTGGGCTTCTTCAGC	58	179 (M), 790 (M/F)	EU687563
SRY	Sex determining region Y	F:TGCATTCATGGTGTGGTCTC R:ATGGCAATTTTTCGGCTTC	58	200	EU687565

Table 1 continued

Gene symbol	Gene name	Primers 5′ to 3′	Annealing temperature, °C	Genomic product size, bp	GenBank accession no
**STS-Y*	Steroid sulfatase (microsomal), isozyme S on Y	F:TGTGTGTTTCTGTCATGGGGATTACATC R:CAGACAATGTTTCCCAGTGACAATTGATTA	58	210 (M/F)	AF133205
TMSB4Y	Thymosin (beta) 4 Y	F:ACCCACCCAGCCTCTTACTT R:GCCTAAGCTGCCAATATCCA	58	246	EU687566
TSPY	Testis-specific Protein Y	F: GAAGTCAGGCACACCAGTGA R: TAAGGCTGCAGTTGTCATGC	58	280	EU687567
UBE1Y	Ubiquitin activating enzyme Y	F:TGGCCAACTCACGGCTGATCCAA R:CTTCTCCACTCACCCTACTTGGG	58	210	EU687568
USP9Y	Ubiquitin-specific protease 9 Y	F:GGTTATGAAATGGTCTCTGC R:CGAGTCTGTCCATCAGGAGTC	58	228	EU687569
**UTY*	Ubiquitously transcribed tetratricopeptide repeat gene, Y-linked	F: CAGCTGTTTTCGGTGATGAG R: GCCTCCTTCTCTTCGGTTG	54	110	
YIR2	Inverted repeat 2 Y	F:AGGGTTGGGCTAAGTCACCT R:ACCTTGGATCCAGACTCACG	58	170	EU687570
ZFY	Zinc finger Y	F:TGAGCTATGCTGACAAAAGGTG R:TCTTTCCCTTGTCTTGCTTGA	58	186	EU687571
ZNF33bY	Zinc Finger protein 33b on Y	F:CCACAGCAAATACAGGAGCA R:GTCTGACTCCTCCCCCTTTC	58	800 (M) 3000 (M/F)	EU687572

* genes identified only from BAC end sequences and not recovered from cDNA selection

These primers were used for PCR amplification from genomic DNA of 5 normal male and 5 normal female horses to confirm their male specificity. All PCR reactions were carried out in 10 µl volume containing 1X buffer (Sigma Aldrich), 0.3 pmol of each primer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$ and 0.25 units JumpStart REDTaq DNA polymerase (Sigma Aldrich) and the products were resolved on 2.0 % agarose gels.

cDNA mapping

Male-specific cDNA sequences, as identified in the previous step, were mapped to the ECAY contig map by STS content analysis. Primers designed from cDNA sequences were used to PCR amplify from the DNA of the 183 Y-specific BAC clones that form the current ECAY contig map (Raudsepp et al. 2004b; Raudsepp and Chowdhary 2008). Additionally, primers from all anonymous sequences that are present both in males and females, were used for STS content mapping on the 12 BAC clones which form the minimum tiling path (MTP) of the horse pseudoautosomal region .

Presence of Open Reading Frames (ORFs)

The sequences were analyzed for the presence of putative open reading frames (ORFs) using Sequencher V 4.7 (GeneCodes Co) and NCBI ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/[12]) software packages.

Screening of male equine BAC libraries by PCR

Primers from those cDNA sequences that were not present on ECAY contig map were used to screen the CHORI-241 male horse BAC library (http://bacpac.chori.org/equine241.htm[13]) by PCR. If BACs corresponding to some cDNA sequences could not be found from the CHORI-241 BAC library, TAMU (http://hbz7.tamu.edu/homelinks/bac_est/bac.htm[14]) and INRA (Milenkovic et al. 2002) male horse BAC libraries were screened. Library screening, growing BAC cultures and BAC DNA isolation was carried out as described earlier (Raudsepp and Chowdhary 2008). Briefly, cDNA primers were used to screen BAC library superpools, plate pools and pools of rows and columns to identify individual BAC clones containing specific cDNA sequences. New BACs were cultured in 2YT media containing 30µl/ml chloramphenicol and plated on LB agar plates which also contained 30µl/ml chloramphenicol. Single colonies were picked and verified once again for identity by PCR with cDNA specific primers, inoculated into 100ml of 2YT (with 30µl/ml chloramphenicol) and grown overnight at 37^{o}C. BAC DNA was isolated by alkaline lysis method (Birnboim and Doly 1979; Birnboim 1983) using Qiagen Midiprep kit (Qiagen) according to the manufacturer's instructions. End sequences of the new BAC clones were used for the development of STS markers, STS content mapping and chromosome walking.

BAC end sequencing

Ends of newly isolated CHORI-241 library BAC clones were sequenced in 10 µl reactions using the standard T7 (5′-GCCGCTAATACGACTCACTATAGGGAGAG-3′) and SP6 (5′-CCGTCGACATTTAGGTGACACTATAG-3′) primers and BigDye chemistry. For TAMU and INRA BAC clones, short T7 (5′-TAATACGACTCACTAAGGG-3′) and M13 (5′-CAGGAAACAGCTATGACC-3′) universal primers were used. Reaction products were purified through Sephadex G-50 Spin-columns (BioMax) and resolved on an ABI-3730 capillary sequencer (PE Applied Biosystems). All BAC end sequences were analyzed for the contents of repeats using RepeatMasker and for the contents of known genes using discontiguous MegaBLAST (Table A1). For the repetitive BAC ends, primers were designed from the overlapping BAC end sequences to internally sequence the BACs with repetitive ends.

STS generation and STS content mapping

STS primers were designed from all non-repetitive BAC end sequences using Primer 3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi[11]) and optimized on male and female horse genomic DNA and DNA of the BAC of sequence origin. The primers were used for STS content analysis on the BACs in the ECAY contig map to incorporate new clones into it. Details about all STS markers generated for the male specific region of ECAY are presented in Table A2.

Chromosome walking

STS primers from the BAC ends that extended into gaps in the map were used to screen CHORI-241 BAC library to identify new clones. After end sequencing, primer design and STS content analysis the new clones were incorporated into map. Such 'chromosome walking' continued until the gap was closed or the gap was flanked by repetitive BAC end sequences which made further walking impossible. In some cases, however, BAC library screening by PCR did not identify any new clones. Since only 50% of CHORI-241 library has been pooled into PCR format, such BAC end sequences were used to design overgo primers for hybridizations on the high density filters containing the entire library.

Overgo primer design and filter hybridization

Filter hybridization was carried out with CHORI-241 BAC library filters using radioactively labeled oligonucleotide primers or overgos. Overgos were designed from non-repetitive end sequences of most outstretching BACs for each gap. The primers were constructed manually or using Overgo Maker program (http://www.genome.wustl.edu/tools/?overgo.html[15]) (Table 2).

Table 2: Overgo primers for EACY BACs.

BACs	Overgo Primers/Probes 5′ to 3′	Contig
005.2A8T7	F:GTGTTATCCAAGCTATGCTTCCTT R:CAACTCTTGTCAGGGCAAGGAAGC	IV
008.3G9T7	F:GCAAGTTAAAGCAAGCAACATGGC R:CTGGCAATAATCAGAAGCCATGTT	III
015.2E9 M13	F: GGTGGGTTATCGAGATCTTCTGGC R: AATGAATGTATGCAATGCCAGAAG	II
016.4C5M13	F:GCCCCCACTTTTGTGGTTTCCTCC R:AAAGAAGAGTGAACACGGAGGAAA	II
017D15T7	F:CCAGCTAGAGCAAAGGACACCTGC R:AAACCCACTGACAAGGGCAGGTGT	I
072G23SP6	F:GATGAGAGAATAGAAGAACAGAGT R:AGAAGATACCCGAGACACTCTGTT	III
090B11SP6	F:GAAGGTAACCTGGTGTGTTCCGCT R:TAATCACTGCTTTATTAGCGGAAC	IV
091.4G10M13	F:CTCCGGTAGGAGAGGAAGAATAGG R:ATATTTAGGGTATTAACCTATTCT	III
110.3H12M13	F:GAGTAGACTCTCAAAGAAGCCAGA R:CCTTGGAACTAACAAATCTGGCTT	V
112.1A9T7	F:GCCCTCAGCCGAAGAGTTGAGAAA R:ACACTTACCGCCAGTGTTTCTCAA	V
118.1A9M13a	F:CTGGGGAATGTCAGTTATTTTTGT R:AAAAAACCAAACCCAAACAAAAAT	III
118.1A9M13b	F:CCATATCCTTATGTGTATCAGCCC R:ATCAGGGCTATTATGTGGGCTGAT	III
159F5SP6	F:GGAGGAAAAACGTAATAAAATCTC R:GAGTGGTCAATCGTATGAGATTTT	II
168I4T7	F:GGCTCACACTTCCTTCCTCTCTGA R:ACTTTCAAAGGGTAGGTCAGAGAG	II
180P20SP6	F:CGTGAGCTGGTGGCTAGTCAGCCC R:TCCATGTAGAATTGCAGGGCTGAC	IV
190M2T7	F:TCACCTTTGAATCATGAAACCCAA R:TTGAGAACTCTTTACCTTGGGTTT	II
331E10SP6	F:GTTAAGGCAGGTTGTCCCCCAGTT R:CCTGCATCATCAGTGAAACTGGGG	IV
C-BWM13	F:GATTCAGAGGCACAGACAGAAACA R:CCCTGCTTTCCAACTCTGTTTCTG	I
D-BWM13	F:GCCAGTCCTGTCAGTGCTCCAATT R:TTATCTGTTTCACCTGAATTGGAG	I

The overgos were labeled individually with ^{32}P dATP and ^{32}P dCTP and the unincorporated nucleotides were removed using Sephadex G-10 spin-columns. Labeled overgo probes were pooled in equal concentrations and hybridized onto the eleven high density filters of the CHORI-241 library using the protocol described by (Gustafson et al. 2003). The filters were washed three times at 42°C for 15 min in 2× SSPE and exposed to film over intensifying screens for four days at –80°C before developing the autoradiograms. The autoradiograms were inspected manually and positive BAC clones were identified. The clones were picked from the BAC library and grown overnight in 96-deep-well plates containing 2YT with antibiotics. BAC clones corresponding to individual overgos were identified by PCR on cell lysates using STS primers from the same BAC end sequences from where overgos were designed. Newly identified BACs were used for further chromosome walking as described above.

Fluorescence in situ hybridization* (FISH) *analysis

FISH was used to confirm the Y-specificity of flow sorted ECAY and all BAC clones. FISH was also instrumental for distinguishing between single copy and multicopy cDNA sequences. Probe labeling, *in situ* hybridization, signal detection and image analysis was carried out according to our detailed protocol (Raudsepp and Chowdhary 2008). Briefly, 1 µg of probe DNA was labeled with biotin and/or digoxigenin using Bio-Nick or Dig-Nick Translation Kit (Roche Molecular Biochemicals) and hybridized to male horse metaphase and/or interphase chromosomes. Biotin labeled probes were detected with avidin-FITC (Vector) and digoxigenin labeled

probes with anti-digoxigenin-rhodamine (Roche Molecular Biochemicals). Images were captured and analyzed with a Zeiss Axioplan2 fluorescence microscope equipped with Isis v 5.2 (MetaSystems GmbH) software.

RESULTS

Isolation of Y-specific ESTs from horse testis

Two rounds of cDNA selection were carried out using horse testis cDNA as a 'driver' and flow sorted Y chromosome or ECAY BAC pools as 'selectors'. From thousands of transcripts present in testis, such selective hybridization allowed isolation of only Y-specific cDNA sequences (Fig. 2). Selected cDNA was cloned and *EcoRI* digestion of 152 randomly chosen clones confirmed the presence of inserts with an average size of ~850bp (Fig. 3). A total of 2,400 clones (864 from BAC pools and 1,536 from flow sorted Y) were picked, sequenced and analyzed. After vector and quality trimming, 1,678 sequences were assembled into 180 contigs and 100 singletons. PCR analysis showed that 321 sequences (19.1%; 30 contigs, 74 singletons) were male specific (Fig. 4). The remaining 1,357 sequences (150 contigs, 26 singletons) amplified both from male and female genomic DNA. Majority of these sequences aligned with known autosomal genes and were removed from further analysis. Fourtyfour non sex-specific cDNA sequences which showed no significant BLAST hits, were analysed by STS content analysis for possible pseudoautosomal origin.

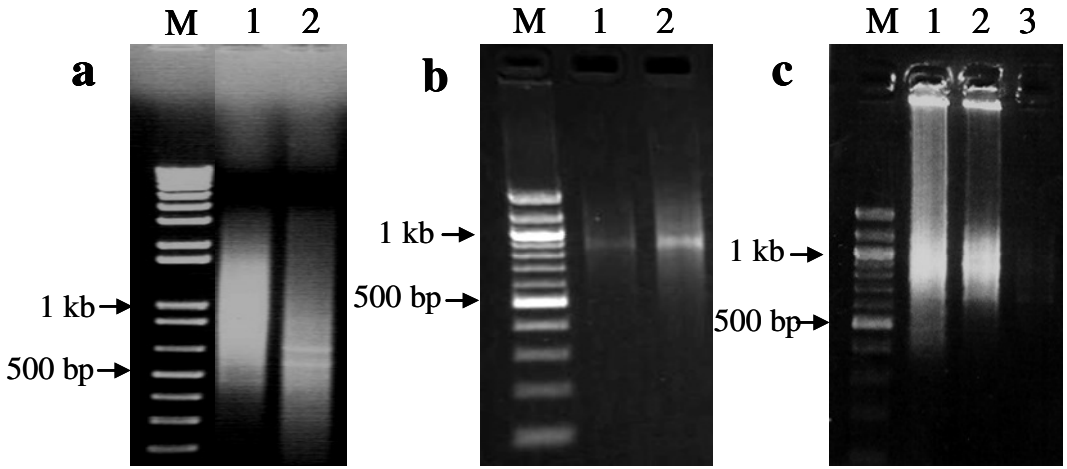

Figure 2: PCR results at different stages of testis cDNA selection. **a.** PCR amplified primary cDNA **b.** PCR amplified primary-selected cDNA, and **c.** PCR amplified secondary selected cDNA, with – 1µl (lane1) and 5µl (lane2) template DNA. M: molecular markers.

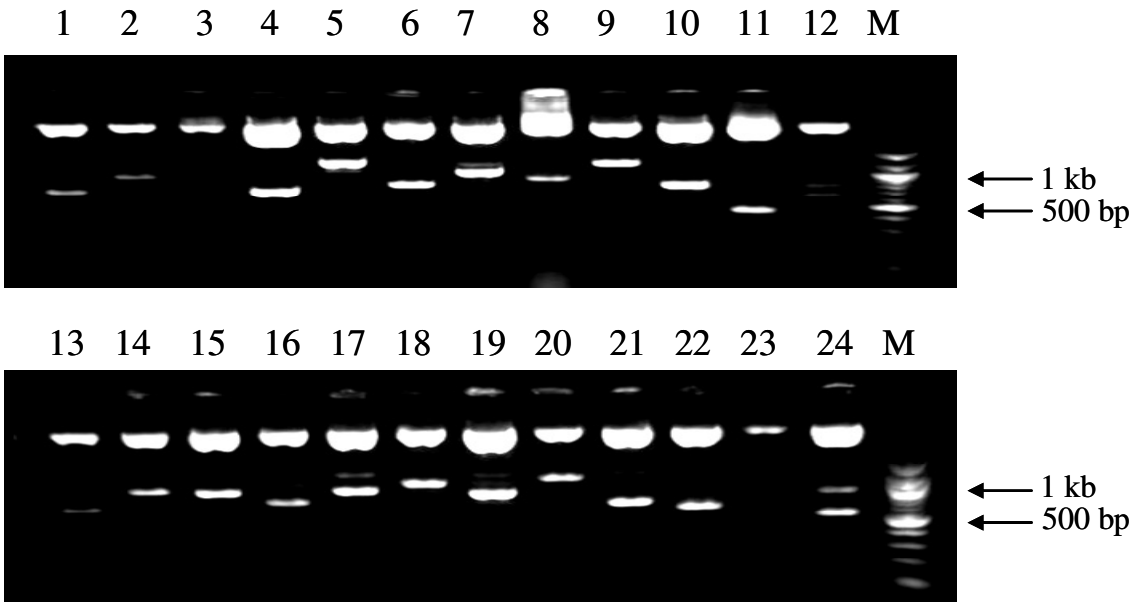

Figure 3: Restriction digestion analysis of cDNA clones. Randomly chosen 24 cDNA clones digested with EcoRI showing inserts of different sizes (lower bands); upper bands represent the Topo TA vector (3.956 kb). M: molecular markers.

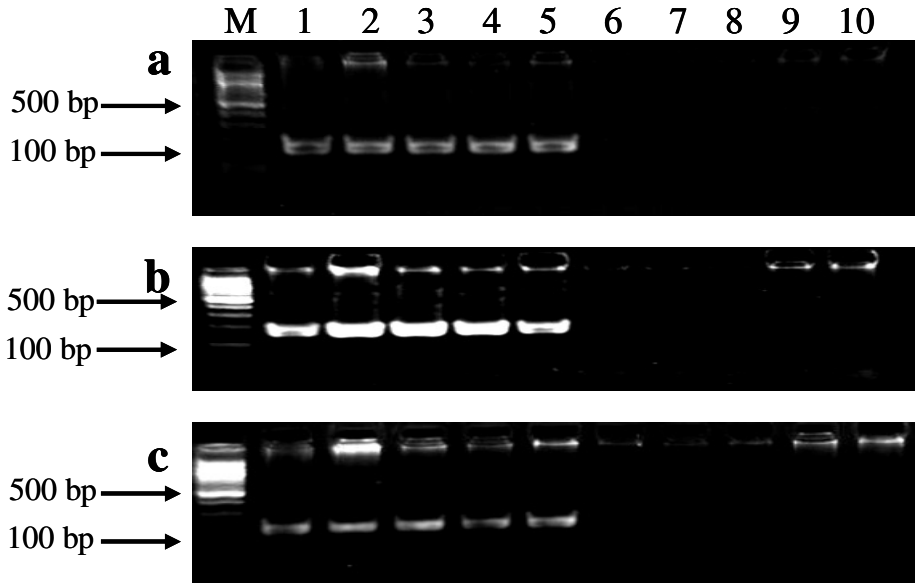

Figure 4: Male-specific PCR amplification. PCR amplification of **a.** *CUL4BY*; **b.** *YIR2*, and **c.** *RFX5Y* from genomic DNA of five male (lanes 1-5) and five female (lanes 6-10) horses. M: molecular markers.

Identification of horse Y chromosome genes and ESTs

BLAST analysis of 321 male-specific cDNA sequences revealed significant (<1e-8) similarity with 13 known mammalian Y chromosome genes *viz*., *DDX3Y, CUL4BY, CYorf15, NLGN4Y, RBMY, KDM5D (alias SMCY), SRY, TMSB4Y, TSPY, UBE1Y, USP9Y, YIR2, ZFY* and 3 autosomal genes, (*viz*., *EIF3C , RPS3A, ZNF33b)* (Table 3). PCR primers for the latter three genes amplified two distinct bands: a higher molecular weight band which was present in both males and females, and a smaller band present only in male DNA (Fig. 5). Sequences of male-specific bands differed from autosomal PCR products by a short internal deletion and it was therefore not possible to design solely male-specific primers.

Ten cDNA sequences were considered as equine specific because BLAST analysis did not show similarity to any mammalian sequences. These were named as *ETSTY1-ETSTY6* and *ETY1-ETY4* where *ETSTY* stands for Equine Testis Specific Transcript on Y and *ETY* for Equine Transcript on Y indicating whether the transcripts had testis-limited or broader expression profiles, respectively (will be discussed in detail in the following chapter).

Finally, three cDNA sequences showed weak similarity (>1e-8) to one mitochondrial (*MT-ND1Y*), one autosomal (*RFX5Y*) and one Y chromosome (*EIF1AY*) gene and their annotation remains therefore tentative.

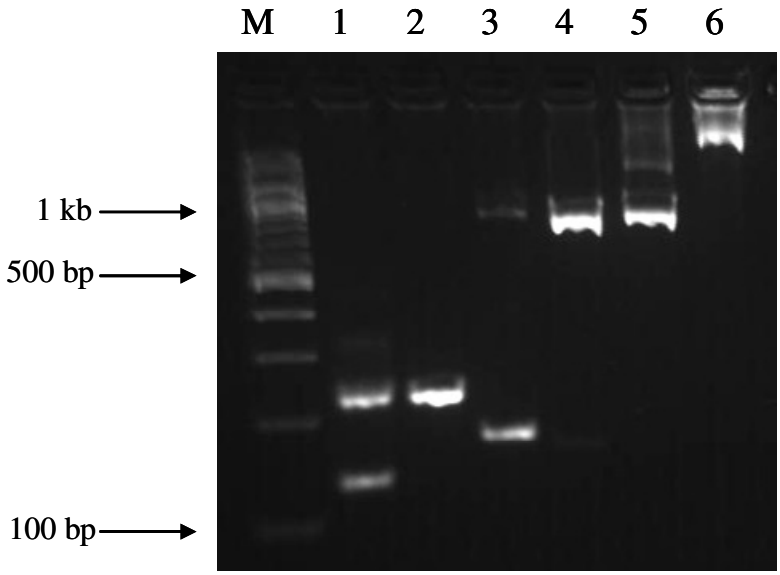

Figure 5: Autosomal and Y-chromosomal PCR amplification. PCR amplification of *EIF3CY* (lanes 1-2), *RPS3AY* (lanes 3-4) and *ZNF33bY* (lanes 5-6) from male (odd numbered lanes) and female (even numbered lanes) genomic DNA. Lower bands are present only in males indicating their Y-specific nature. M: molecular markers.

Table 3: Summary information about all genes and ESTs mapped to ECAY.

Gene Symbol	Copy number		Gene/EST category	Presence of ORF	No. of cDNA clones isolated	Known homologs in other mammals
	BAC	cDNA				
**AMELY*	Single copy	-	X degenerate	No	0	Human, pig, cattle *AMELY*
CUL4BY	Multicopy	Heterochromatin	X degenerate, multicopy	Yes	1	Cat, dog *CUL4BY*
CYorf15	Single copy	-	X-degenerate	Yes	2	Human *CXorf15*, cat *CYorf15*
DDX3Y	Single copy	-	X degenerate	Yes	2	Human, mouse, cat, pig *DDX3Y*
EIF1AY	Single copy	-	X-degenerate	No	2	Human *EIF1AY*
**EIF2s3Y*	Single copy	-	X degenerate	No	0	Mouse, pig, cat *EIF2s3Y*
EIF3CY	Single copy	-	Autosomal	Yes	85	Cattle, Dog, Human *EIF3C*
ETSTY1	Multicopy	Multicopy	Novel, multicopy	No	2	No significant similarity
ETSTY2	Multicopy	Multicopy	Novel, multicopy	No	7	No significant similarity
ETSTY3	Multicopy	Multicopy	Novel, multicopy	No	12	No significant similarity
ETSTY4	Multicopy	Multicopy	Novel, multicopy	No	9	No significant similarity
ETSTY5	Multicopy	Multicopy	Novel, multicopy	No	3	No significant similarity
ETSTY6		Heterochromatin	Novel, multicopy	No	2	No significant similarity
ETY1	Multicopy	Multicopy	Novel, multicopy	No	1	No significant similarity
ETY2	Single copy	-	Novel	No	1	No significant similarity
ETY3		Heterochromatin	Novel, multicopy	No	1	No significant similarity
ETY4	Multicopy	Multicopy	Novel, multicopy	No	1	No significant similarity
**KAL1Y*	Single copy	-	X degenerate	No	0	Pseudogene in human
KDM5D (SMCY)	Single copy	-	X degenerate	Yes	3	Horse, dog, human *SMCY*
**MAP3K7IP3 Y*	Single copy	-	X degenerate	No	0	Human *MAP3K7IP3* on X chromosome
MT-ND1Y	Single copy	-	Mitochondrial	No	1	Human *MT-ND1* pseudogene on X chromosome, NADH dehydrogenase 1 (MTND1) pseudogene on HSA1

Table 3 continued

Gene Symbol	Copy number		Gene/transcript category	Presence of ORF	No. of cDNA clones isolated	Known homologs in other mammals
	BAC	cDNA				
NLGN4Y	Single copy	-	X degenerate	No	1	Human *NLGN4Y*, Macaca mulatta, chimp *NLGN4X*
RBMY	Multicopy	Multicopy	X degenerate, multicopy	No	1	Human, mouse *RBMY*, Human, chimp *RBMX*
RFX5Y	Single copy	-	Autosomal	No	1	Human *RFX5*
RPS3AY	Single copy	-	Autosomal	Yes	1	Human *RPS3A*
SRY	Multicopy	Single copy	X degenerate	Yes	43	Human, mouse, cat, pig, cattle *SRY*
**STS-Y*	Single copy	-	X degenerate	No	0	Mouse *Stsy* pseudoautosomal
**TBL1Y*	Single copy	-	X degenerate	No	0	Human *TBL1Y*
TMSB4Y	Single copy	-	X degenerate	No	1	Human, mouse *TMSB4Y*, cattle, human *TMSB4X*
TSPY	Multicopy	Multicopy	X degenerate, multicopy	Yes	12	Human, chimp, cat, cattle, pig *TSPY*
UBE1Y	Multicopy	Multicopy	X degenerate, multicopy	Yes	54	Human *UBE1*, mouse, pig, cat *UBE1Y*
USP9Y	Single copy	-	X degenerate	Yes	3	Human, chimp, rat, mouse *USP9Y*
**UTY*	Single copy	-	X degenerate	No	0	Human, mouse, pig, cat *UTY*
YIR2	Multicopy	Multicopy	Inverted repeat, multicopy	No	1	Human chromosome Y palindromes P1, P2, P3 and inverted repeat IR2 (P1-P2-P3-IR2@) on chromosome Y
ZFY	Single copy	-	X degenerate	No	3	Human gorilla, chimp *ZFY*, human *ZFX*
ZNF33bY		Heterochromatin	Autosomal, multicopy	No	64	Equus caballus similar to zinc finger protein 33b (*ZNF33b*)

* genes identified from BES only

Altogether, the analysis of all male specific cDNA sequences identified 29 known or novel equine Y chromosome genes and ESTs. Four genes, *viz., SRY, CUL4BY*, *RPS3AY* and *ZNF33bY* were captured only from the experiments with flow sorted ECAY, 12 genes using ECAY BAC pools and the remaining 13 ESTs were identified from both types of experiments (Table 3).

Relative abundance of Y-linked transcripts

The most abundant male-specific transcripts were *EIF3CY, ZNF33bY, UBE1Y* and *SRY* represented by 85, 64, 54 and 43 cDNA sequences, respectively (Table 3). It is likely, that the number of *EIF3CY* and *ZNF33bY* transcripts is inflated because sequence analysis could not clearly distinguish between the sequences originating from Y-specific and autosomal transcripts of the two genes. Therefore, the most highly transcribed Y chromosome genes in adult horse testis are probably *UBE1Y* and *SRY*. In contrast, only one transcript was found for *CUL4BY, ETY1-ETY4, MT-ND1Y, NLGN4Y, RBMY, RFX5Y, RPS3AY, TMSB4Y,* and *YIR2* (Table 3).

Identification of open reading frames (ORFs)

The presence of ORFs was identified in 10 sequences, *viz., CUL4BY, CYorf15, DDX3Y, EIF3CY, RPS3AY, KDM5D (alias SMCY), SRY, TSPY, UBE1Y* and *USP9Y* (Table 3). Because many captured cDNA sequences were short the final count for the presence or absence of ORFs can be done only after full-length cDNA sequences are available for all ECAY ESTs.

cDNA mapping on horse Y chromosome

Twenty-six Y-linked genes and ESTs were mapped to ECAY contig map (Raudsepp et al. 2004b; Raudsepp et al. 2008b) by STS content analysis. The location of six genes (*DDX3Y, KDM5D (alias SMCY), SRY, TSPY, USP9Y, ZFY*) confirmed our earlier results (Raudsepp et al. 2004b) while the rest were new map assignments. Sequences of three ESTs (*ZNF33bY*, *ETY3* and *ETSTY6*) were not found from the existing contig map and were used to isolate new clones from CHORI-241 library. From the total of 29 Y-specific genes and ESTs 12 mapped to the multicopy region, 14 to the remaining MSY, and *ZNF33bY*, *ETY3* and *ETSTY6* to the heterochromatic region (Fig. 6). The linear order of all single-copy genes was determined. In the multicopy region it was possible to order only three genes, *viz*., *RBMY*, *SRY* and *YIR2*, because they are located in the same BAC clone and a complete sequence of this clone is available at NCBI Entrez Nucleotide, Accession AC214740. The order of the remaining genes and ESTs in multicopy and heterochromatic regions remain tentative.

An additional seven genes viz., *KAL1Y*, *TBL1Y*, *UTY*, *STSY, EIF2s3Y, MAP3K7IP3Y* and *AMELY* were identified only from BAC end sequence (BES) analysis and not found from cDNA selection experiments. Therefore, the current gene map of horse MSY consists of 36 genes and ESTs (Fig. 6).

Primers for 44 anonymous cDNA sequences that PCR amplified equally from male and female genomic DNA, were used for STS content analysis on the minimum tiling path (MTP) of horse PAR (Raudsepp and Chowdhary 2008). Three transcripts were initially assigned to the horse pseudoautosomal boundary (PAB) and one to ECAX,

adjacent to PAB. However, recent re-analysis of these sequences with RepeatMasker showed that all four cDNA sequences are completely repetitive and were hence eliminated from further study. None of the cDNA sequences analyzed in this research showed significant similarity to any known equine or other mammalian pseudoautosomal genes.

Generation of a comprehensive map for ECAY

The initial contig map of ECAY comprised of 73 BACs, 126 STS markers and 8 genes (*KDM5D, SRY, TSPY, UTY, USP9Y, DDX3Y, ZFY,* and *AMELY)* that were arranged into seven BAC contigs in MSY (Raudsepp et al. 2004b). Each contig contained 3-13 overlapping BACs with the exception of the multicopy region which had 27 clones.

Since then the map has been considerably expanded. cDNA selection in this study identified a number of new genes/ESTs which led to isolation of many new BAC clones and generation of STS markers for chromosome walking. If BAC ends were repetitive, internal BAC sequences were used to walk through the repeat region. An additional 39 new clones were found by overgo hybridizations to the CHORI-241 library high-density filters. Altogether, chromosome walking and BAC discovery identified 124 new clones that added one new contig to the map (contig III with 27 BACs Figs. 6, A1) and were instrumental in closing two gaps. The final map spans the entire MSY and extends from the pseudoautosomal boundary to the heterochromatic region and even includes a small part of the latter. The map consists of 197 BAC clones which are

arranged into 5 contigs. Majority of the BAC clones (143) originate from CHORI-241 male BAC library, though 54 BAC clones were incorporated from TAMU (42) and INRA (12) libraries. The use of Y chromosome from three different individuals did not affect the assembly of the contig map because ECAY exhibits overall very low genetic variability (Lindgren et al. 2004; Wallner et al. 2004). The present map contains 282 STSs and 36 linearly ordered genes. A contig-wise summary of BES, STSs and genes present in horse MSY is provided in Table A3.

Despite this progress, the map still contains four gaps (Fig. A1) which could not be closed because of the presence of massive repetitive sequences flanking the gaps. Co-hybridization of BACs from either side of the gaps showed that GAP1 and GAP4 are the smallest, not larger than 2-3 BAC insert sizes or ~ 500 kb, while GAP2 and GAP3 are almost two times larger, about the size of 5-6 BAC inserts or ~ 1 megabase (Mb) each. Altogether, the four gaps cover at least 3 Mb of the horse MSY (Fig. A1).

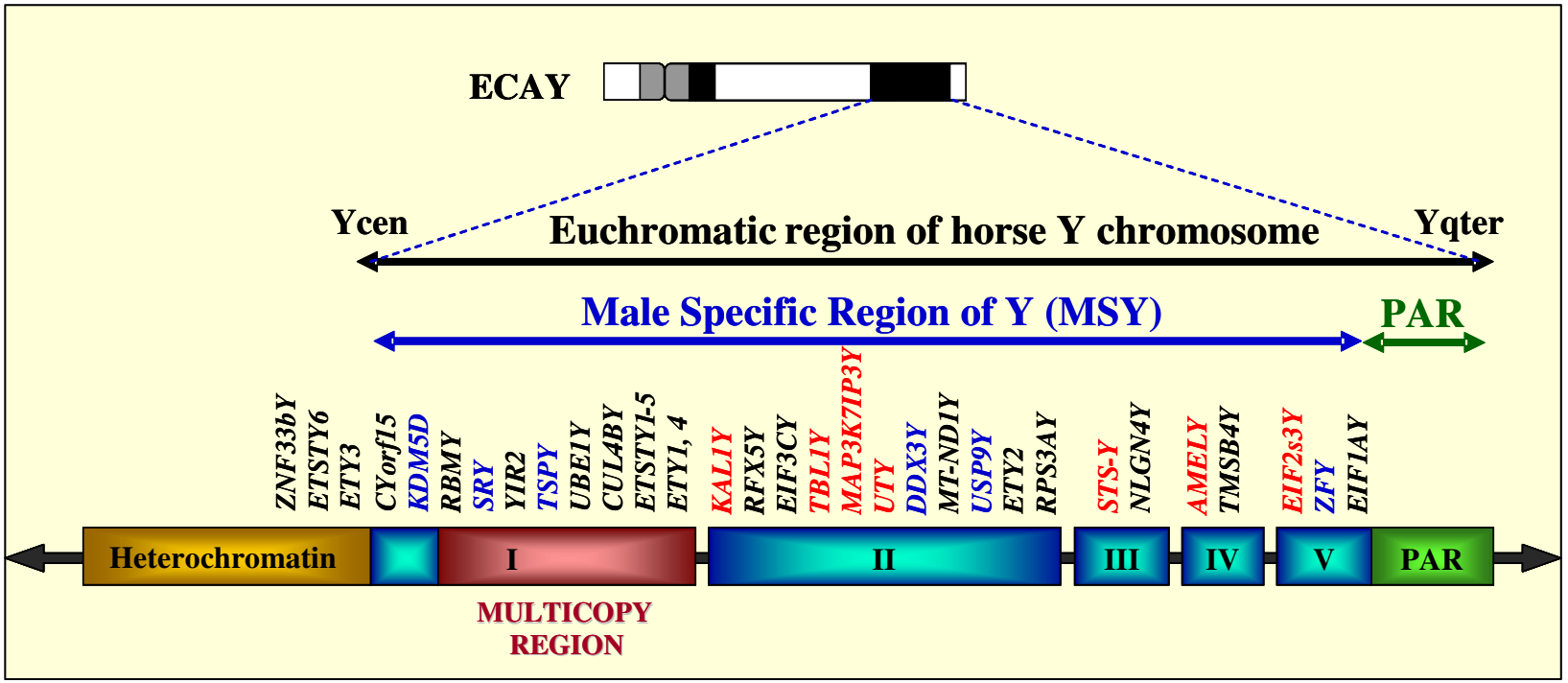

Figure 6: Schematic representation of ECAY heterochromatic region, MSY and PAR showing the five MSY BAC contigs and location of 36 genes and ESTs. Gene symbols in black – newly added from cDNA selection; blue – previously mapped (Raudsepp et al. 2004) and red – identified from BES only.

FISH analysis and cDNA copy number evaluation

FISH confirmed Y specificity of the flow sorted ECAY, the six BAC pools and all newly isolated BAC clones (Fig. 7a, b). FISH also allowed to distinguish between single-copy and multicopy BAC clones by showing the presence of one or more signals on metaphase or interphase chromosomes. However, FISH with multicopy BAC clones does not reveal whether the short cDNA sequences contained in these BACs are also multicopy. Therefore, direct FISH with short cDNA clones was used to evaluate copy number of individual genes and ESTs. The underlying rationale of this approach is the limited sensitivity of the FISH technique (Trask et al. 1993; Trask 2002) due to which signals from short (600-900 bp) cDNA sequences are visible only when these sequences are present in multiple copies. cDNA FISH showed that 15 genes and ESTs are present in multiple copies on the Y chromosome (Fig. 7c, d; Table 3, 4). Not coincidentally, 12 markers are located in BAC clones from multicopy region (Fig. 6). The only exception was *SRY* cDNA which produced no detectable FISH signals though *SRY* co-localizes with three multicopy genes – *RBMY, TSPY* and *YIR2* – in the same multicopy BAC clone (Fig. A1, Table 3). It is likely that horse *SRY* is a single copy gene but is inserted between multicopy sequences.

cDNA from *ETY3*, *ETSTY6* and *ZNF33bY* hybridized to ECAY and ECAXq21 heterochromatin. Heterochromatic signals were obtained also with *CUL4BY* cDNA, though STS content analysis mapped this gene unequivocally to the multicopy region in Contig I (Fig. 6).

cDNA FISH allowed mainly to distinguish between single copy and multicopy sequences (Fig. 7), though in a few cases the signal intensity indicated also quantitative differences between the markers. For example, *RBMY* cDNA gave consistently brighter signals than *ETSTY2* cDNA reflecting likely differences in copy numbers (Fig. 7f, g).

DISCUSSION

cDNA selection and ECAY gene map

cDNA selection is a well established method to isolate expressed sequences from genomic regions of interest (Chen-Liu et al. 1995; Guan et al. 1996). In the case of the Y chromosome where gene discovery through direct sequencing or comparative mapping is complicated, cDNA selection has been an efficient way to obtain a broad representative sampling of expressed sequences from MSY and PAR in humans (Lahn and Page 1997; Makrinou et al. 2001) and more recently in carnivores (Murphy et al. 2006; Pearks Wilkerson et al. 2008). The present study shows the efficacy of the method also in horse where 29 genes and ESTs were identified of which 23 were new for the horse Y chromosome.

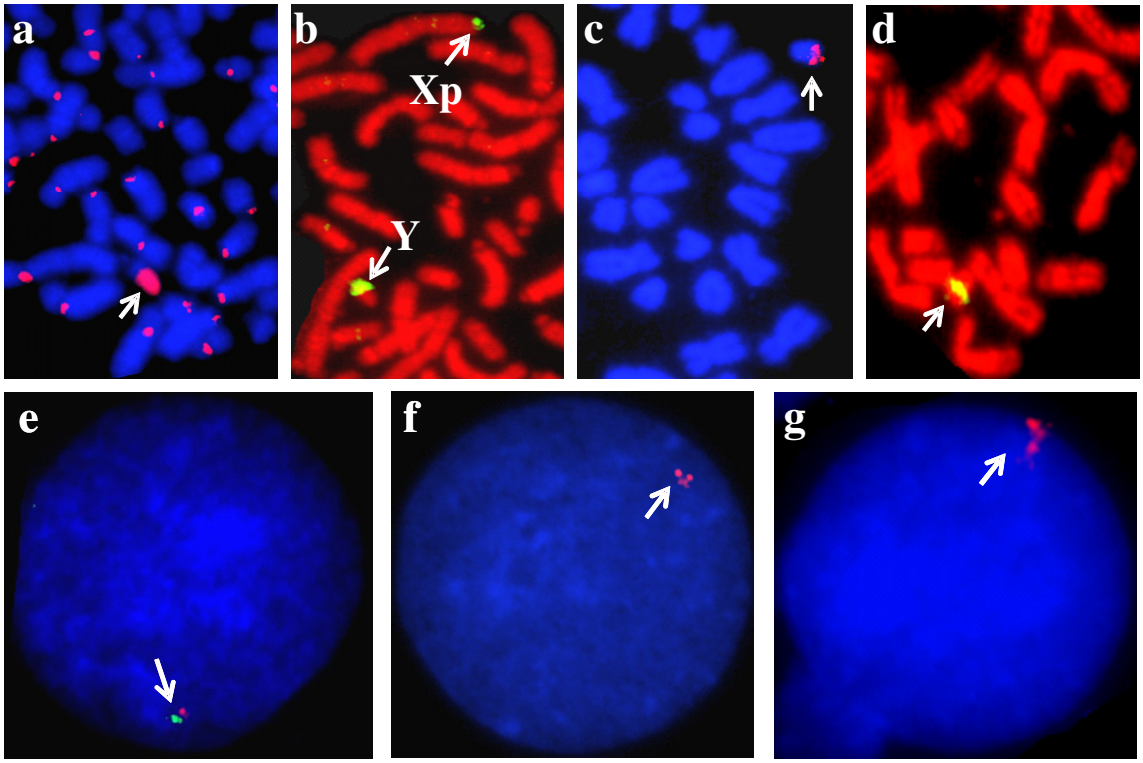

Figure 7: Fluorescence *in situ* hybridization (FISH) using ECAY probes. FISH showing signals (arrows) on horse metaphase (a-d) and interphase (e-g) chromosomes using **a.** dig-labeled flow-sorted Y chromosome, **b.** biotin labeled pool of 33 ECAY BACs. An additional signal is seen on Xpter because the pool contains a few clones which are duplicated between MSY and PAR (Raudsepp & Chowdhary 2008), **c.** dig-labeled *ETSTY3 cDNA,* **d.** biotin labeled *UBE1Y* cDNA, **e.** co-hybridization of BACs 124.3G9 (red) and 83H5 (green) containing single copy genes *EIF2s3Y* and *AMELY*, respectively, **f.** dig-labeled *ETSTY2* cDNA, and **g.** dig-labeled *RBMY* cDNA.

Although the search for ECAY genes using flow sorted Y chromosome and 183 Y specific BAC clones was systematic, some previously mapped genes (Raudsepp et al. 2004b; Raudsepp et al. 2008b) escaped detection. For example, *AMELY* was not found because it is expressed exclusively in tooth enamel and not in testis (Salido et al. 1992). *STS-Y* and *KAL1Y* escaped discovery probably because, like in humans, they might be non-transcribed pseudogenes (Skaletsky et al. 2003). *UTY, MAP3K7IIP3Y*, *EIF2s3Y* and *TBL1Y* were not recovered probably because they are not transcribed in sufficient amounts in horse testis to be captured by the cDNA selection method. A few more pseudogenes and low-abundance transcripts might have escaped detection. For example, two known mammalian Y-linked genes – *HSFY* and *RPS4Y* – have not yet been found in horses. It is also possible that these genes have been lost from ECAY.

Surprisingly, we did not recover any known equine (Raudsepp and Chowdhary 2008) or other mammalian PAR genes though initial cDNA capture for human Y chromosome identified all 9 PAR genes known at that time (Lahn and Page 1997). It is possible that PAR gene expression level in horse testis is lower than in humans and these transcripts escaped detection in the present study. Nonetheless, the cDNA selection enriched horse Y chromosome gene map with 23 new genes and ESTs.

Comprehensive contig map of ECAY

Using the initial ECAY contig map (Raudsepp et al. 2004b) as an important foundation, the construction of a comprehensive BAC contig map for horse MSY was completed in this study. The map covers most of the MSY and provides a highly

redundant tiling path for the sequencing of the horse Y chromosome (Fig. A1). The most proximal contig connects MSY to the heterochromatic portion of ECAY and four most distal clones extend over the peudoautosomal boundary and join MSY to the PAR (Raudsepp and Chowdhary 2008). The minimum tiling path (MTP) over MSY comprises of 59 BAC clones of which 44 are single-copy and 15 multicopy. The sequence-ready contig map of the horse Y chromosome is the only such resource for a Y chromosome among non-primate mammals.

One of the goals of the present study was also to close gaps present in the intitial MSY BAC contig map (Raudsepp et al. 2004). For this several different approaches were used: i) cDNA selection from horse testis to enrich the map with randomly spaced genes and ESTs, ii) filter hybridizations with overgo primers to identify new Y-specific BACs and iii) sequencing of BAC ends and internal sequences to develop more STSs for chromosome walking. This resulted in closing two of the six gaps in the previous MSY map (Raudsepp et al. 2004b) (Fig. 6, A1). It is noteworthy that majority of the isolated cDNA clones mapped to the existing BACs and none to the gaps. This indicates that most likely the gaps do not contain any expressed sequences. It is possible that the gaps are enriched with repetitive sequences and have therefore largely escaped cloning into libraries. The gaps might also demarcate transition from euchromatic sequences into small interstitial heterochromatic islands. Such islands have been recently discovered from human MSY and are often flanked by segmental duplications (Kirsch et al. 2008). Whether ECAY contains similar interstitial heterochromatic islands needs further investigation.

Given that the CHORI-241 library average insert size is 171 kb (http://bacpac.chori.org/equine241.htm[13]), that the 59 MTP clones are overlapping by ~20% and that the four gaps count for ~ 3 Mb, the size of equine MSY is estimated to be approximately 11 Mb. Together with PAR which is ~1.8 Mb (Raudsepp and Chowdhary 2008) the size of the horse Y chromosome euchromatin is roughly 13 Mb. This estimate is close to the ~15 Mb proposed earlier (Raudsepp et al. 2004b) and will be validated once the complete sequence of the euchromatic ECAY becomes available.

Horse Y chromosome gene catalogue

Traditionally MSY sequences are classified as X degenerate or ampliconic (Skaletsky et al. 2003) reflecting their evolutionary origin and copy number. The two classes, however, are not always clearly distinguished. The best example is *TSPY* which represents both sequence categories. It has an ancestral single-copy and ubiquitously expressed homolog on the X chromosome (Delbridge et al. 2004), but has been amplified and has acquired testis-limited expression on the Y chromosome in human (Skaletsky et al. 2003), cat (Murphy et al. 2006), horse (Raudsepp et al. 2004b; this study), cattle, goat, sheep (Vogel et al. 1997a; Vogel et al. 1997b) and rat (Dechend et al. 1998). In contrast, murine *TSPY* is a single copy gene but has lost its function (Schubert et al. 2000b).

According to their evolutionary origin, the 36 horse MSY genes and ESTs identified and mapped in this study were classified as X degenerate, multicopy and acquired sequences (Fig. 6, Table 3). The former originate from ancestral proto-sex

chromosomes, have a gametologue on the X chromosome and are often shared between mammalian Y chromosomes (Graves 2006; Delbridge and Graves 2007; Waters et al. 2007). Most of the multicopy genes have no detectable homologs in other species and are most likely Y-borne in horses. In contrast, acquired sequences have arrived to the Y chromosome from other genomic locations.

X degenerate sequences

The core set of horse X degenerate genes resembles, with minor variations, that of other mammals (Quilter et al. 2002; Rohozinski et al. 2002; Skaletsky et al. 2003; Raudsepp et al. 2004b; Murphy et al. 2006; Pearks Wilkerson et al. 2008). Like in other species, the majority of equine X degenerate genes are single copy sequences. Horse-specific variations involve possible loss of some genes, *viz., HSFY* and *RPS4Y* and structural-functional changes in a few others. Besides *TSPY*, three more X degenerate genes - *RBMY*, *CUL4BY* and *UBE1Y* - have been amplified during ECAY evolution. *RBMY* is a multicopy gene in human and mouse (Mazeyrat et al. 1999; Skaletsky et al. 2003), multiple copies of *CUL4BY* have been found also from carnivore Y chromosome (Murphy et al. 2006; Pearks Wilkerson et al. 2008), while amplification of *UBE1Y* seems to be a horse specific feature. This gene has been lost from human Y chromosome (Lahn et al. 2001; Skaletsky et al. 2003) but is present as a single-copy Y-linked sequence in cat (Murphy et al. 2006), pig (Quilter et al. 2002) and mouse (Mitchell et al. 1991; Levy et al. 2000). Functional importance of *UBE1Y* in horse testis is reflected by the higher than average number of recovered cDNA clones (54 cDNAs, Table 3).

UBE1Y transcripts were so abundant that in some experiments, in order to efficiently recover single-copy sequences, the initial testis cDNA was pre-annealed with *UBE1Y* cDNA. However, given that similarity between *UBE1Y* and *UBE1X* coding sequences is high (95-98%), the recovered pool of cDNAs could originate from both gametologues and thus, artificially inflate the numbers.

The most intriguing X degenerate gene, however, seems to be the male sex-determining region on Y – *SRY*. Considering the known function of *SRY* in sex determination at early stages of mammalian embryonic development (Wilhelm et al. 2007b), it comes as a surprise that *SRY* is transcribed at high levels from adult horse testis (43 cDNA clones recovered, Table 3). One explanation for elevated transcription levels could be gene amplification, especially in the light that *SRY* is a multicopy gene in rabbit (Geraldes and Ferrand 2006), rat (Turner et al. 2007) and cat (Pearks Wilkerson et al. 2008). Mapping *SRY* to a multicopy BAC clone, very close to three other multicopy transcripts (*RBMY, TSPY* and *YIR2)* (Fig. 6) indicates a possible multicopy nature of horse *SRY*. However, *SRY* cDNA FISH showed no hybridization signal and we infer that equine *SRY* is a single copy gene which is located in the middle of amplified sequences. Finally, horse *SRY* exhibits one more unusual feature – the sequence of the single 1420 bp coding exon (Hasegawa et al. 1999, this study) contains a ~20 bp LTR repeat which was discovered during cDNA selection. It appeared that *SRY* transcripts could be recovered only when the amount of Cot1 DNA used for pre-annealing testis cDNA was reduced from 7.5X to 2X. Sequence analysis indicates that intra-exonic LTR or simple repeats are present also in mouse, rat, rabbit, dog and donkey *SRY* (Ensembl,

http://www.ensembl.org/index.html[3]; NCBI Entrez Nucleotide, http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide[16]). The functional significance of these repeat sequences is yet to be determined.

Multicopy sequences – novel equine-specific Y-borne transcripts

The presence of species specific, Y-borne or acquired multicopy sequences seems to be one of the unique and characteristic features of all Y chromosomes studied so far – from mammals to the fruit fly (Skaletsky et al. 2003; Raudsepp et al. 2004b; Gvozdev et al. 2005; Toure et al. 2005; Murphy et al. 2006; Delbridge and Graves 2007). The horse MSY is no exception as we recovered 15 different multicopy cDNA sequences of which 9 were considered novel, equine-specific and Y-borne because none showed similarity to known autosomal or X-linked sequences. The remaining 6 ESTs showed sequence similarity to putative human, mouse and cat Y orthologs and human autosome (Table 3). It is worth mentioning that out of 10 novel ESTs found in this study only one (*ETY2*) was a single copy sequence.

In humans where MSY sequence is available, ampliconic regions are defined as amplified segments of euchromatic sequences that exhibit as much as 99.9% identity over 10-100 kilobases to other MSY sequences (Skaletsky et al. 2003). In species where Y chromosome sequence information is not available, ampliconic nature of sequences is determined indirectly by genotyping on radiation hybrid panel (Murphy et al. 2006) or by FISH (Raudsepp et al. 2004b; Murphy et al. 2006). In the present study multicopy

status of all 15 sequences was determined solely by FISH and therefore exact copy number and rates of sequence similarities between amplicons remain unknown.

Most of the novel cDNA sequences mapped to a region in the distal part of MSY Contig I and two sequences to the heterochromatic region (Fig. 6). Such regional localization of multicopy sequences, as far as known, is unique to the horse Y chromosome. Human ampliconic sequences are distributed among at least 5 distinct regions and the sequence with the highest copy number (35 copies) is *TSPY* (Skaletsky et al. 2003). In cat and mouse high numbers of amplified and expressed sequences are dispersed over the entire long arm of the Y chromosome and in mouse some testis-specific gene families (*Ssty, Asty*) each are present in as many as 65-100 copies (Toure et al. 2004a; Toure et al. 2004b; Ellis et al. 2005; Murphy et al. 2006). Copy number difference between different multicopy cDNA sequences (based on variable intensity and numbers of FISH signals) was noted also in this study (Fig. 7 d, e, f). These observations, however, need validation and refinement by quantitative PCR and/or complete sequencing of horse MSY.

Notably, we did not detect ORFs in any of the novel multicopy transcripts (Table 3). It is possible that like in human and cats, the multicopy region of horse MSY consists of tandemly repeated transcription clusters that lack strong protein-coding evidence (Skaletsky et al. 2003; Murphy et al. 2006). Protein coding potential of these novel multicopy sequences will be re-evaluated once full length cDNA sequences are available.

Acquired sequences from autosomes and the mitochondrial genome

Transposition and retroposition of genes and mRNA sequences from autosomes to the Y chromosome is another characteristic feature of Y chromosome evolution (Gvozdev et al. 2005; Delbridge and Graves 2007) and has been shown for human *DAZ* and *CDY* gene families (Saxena et al. 2000; Dorus et al. 2003; Skaletsky et al. 2003), mouse *RhoA* genes (Boettger-Tong et al. 1998) and cat *FLJ36031Ya*, *TETY1* sequences (Murphy et al. 2006). Such gene traffic, however, is species or group specific because in different species different autosomal genes have acquired Y-linked counterparts. This is consistent with the results of the present study showing that none of the autosome-derived genes on horse MSY (Fig. 5) are Y-linked in other mammals (Table 3). Identification of a tentative Y-linked homolog for regulatory factor X, 5 on Y (*RFX5Y*) is of particular interest. Autosomal *RFX5* is a part of conserved upstream elements in the major histocompatibility complex (MHC) class-II promoter (Lochamy et al. 2007). It is, therefore, possible that the horse Y-linked counterpart is associated with the minor histocompatibility complex and Y chromosome encoded H-Y antigens (Rosinski et al. 2008), but this conjecture needs verification in future studies.

Lastly, one Y-linked transcript is probably a mitochondrial insertion showing tentative similarity (>1e-8) to *MT-ND1*. Though nuclear sequences of mitochondrial origin (NUMTs) have not yet been studied in the horse genome, the discovery of one from the Y chromosome is consistent with the findings in human where HSAY is highly susceptible for colonization by NUMTs (Ricchetti et al. 2004). An overview of all ECAY genes that have orthologs in other mammalian species is presented in Table 4.

Table 4: Comparative analysis of ECAY orthologs on the Y chromosomes of other mammalian species.

Horse gene /transcript	ECAY	HSAY	MMUY	FCAY	Other Mammalian Y orthologs	References
AMELY	SC	SC	not found	SC	Chimpanzee, pig, cattle SC	Skaletski et al. 2003; Pearks Wilkerson et al. 2008; Kuroki et al. 2006; Quilter et al. 2002; Liu et al. 2002
CUL4BY	MC	not found	not found	MC	Dog MC	Murphy et al. 2006; Pearks Wilkerson et al. 2008
CYorf15	SC	SC, 2 isoforms	not found	MC	Chimpanzee 2 isoforms	Skaletski et al. 2003; Murphy et al. 2006; Pearks Wilkerson et al. 2008; Kuroki et al. 2006
DDX3Y	SC	SC	SC	SC	Chimpanzee, rat, pig, cattle SC	Skaletski et al. 2003; Pearks Wilkerson et al. 2008; Kuroki et al. 2006; Quilter et al. 2002; Liu et al. 2002
EIF1AY	SC	SC	not found	SC	Chimpanzee SC	Skaletski et al. 2003; Pearks Wilkerson et al. 2008; Kuroki et al. 2006
EIF2s3Y	SC	not found	SC	SC	Pig SC	Pearks Wilkerson et al. 2008, Quilter et al. 2002
EIF3CY	SC	not found	not found	not found	not found	
KAL1Y	SC	SC, pseudogene	not found	not found	Pig, cattle PAR	Skaletski et al. 2003; Quilter et al. 2002; Das et al. unpublished
KDM5D (SMCY)	SC	SC	SC	SC	Chimpanzee, dog, cattle, pig SC	Skaletski et al. 2003; Pearks Wilkerson et al. 2008; Kuroki et al. 2006; Liu et al. 2002; Quilter et al. 2002
MAP3K7IP3Y	SC	not found, X-linked	not found	not found	not found	

Table 4 continued

Horse gene /transcript	**ECAY**	**HSAY**	**MMUY**	**FCAY**	**Other Mammalian Y orthologs**	**References**
MT-ND1Y	SC	SC, pseudogene,	not found	not found	not found	Skaletski et al. 2003
NLGN4Y	SC	SC	SC	not found	Chimpanzee SC, cattle PAR	Skaletski et al. 2003; Kuroki et al. 2006; Das et al. unpublished
RBMY	MC	MC	MC	not found	not found	Skaletski et al. 2003; Toure et al. 2004
RFX5Y	SC	not found	not found	not found	not found	
RPS3AY	SC	not found	not found	not found	not found	
SRY	SC	SC	SC	MC	Chimpanzee, cattle, pig SC, rat MC	Skaletski et al. 2003; Murphy et al. 2006; Pearks Wilkerson et al. 2008; Kuroki et al. 2006; Quilter 2002; Moore et al. 2001; Turner et al. 2007
STS-Y	SC	pseudogene	PAR	not found	Chimpanzee pseudogene, cattle, pig, dog PAR	Skaletski et al. 2003; Kuroki et al. 2006; Perry et al. 2001; Liu and de Leon 2004
TBL1Y	SC	SC	not found	not found	Chimpanzee SC, cattle PAR	Skaletski et al. 2003; Kuroki et al. 2006; Van Laere et al. 2008
TMSB4Y	SC	SC	not found	not found	Chimpanzee SC	Skaletski et al. 2003; Kuroki et al. 2006
TSPY	MC	MC	pseudogene	MC	Chimpanzee, rat, cattle, goat, pig MC	Skaletski et al. 2003; Mazeyrat and Mitchell 1998; Murphy et al. 2006; Pearks Wilkerson et al. 2008; Dechend et al. 1998; Kuroki et al. 2006; Quilter et al. 2002

Table 4 continued

Horse gene /transcript	**ECAY**	**HSAY**	**MMUY**	**FCAY**	**Other Mammalian Y orthologs**	**References**
UBE1Y	MC	not found	SC	2 copies	Rat SC, pig SC	Murphy et al. 2006; Pearks Wilkerson et al. 2008; Quilter et al. 2002
USP9Y	SC	SC	SC	SC	Chimpanzee, rat, pig SC	Skaletski et al. 2003; Pearks Wilkerson et al. 2008; Kuroki et al. 2006; Quilter et al. 2002
UTY	SC	SC	SC	SC	Chimpanzee, pig SC	Skaletski et al. 2003; Pearks Wilkerson et al. 2008; Kuroki et al. 2006; Quilter et al. 2002
YIR2	SC	palindromes P1, P2, P3 and inverted repeat IR2	not found	not found	not found	Skaletski et al. 2003
ZFY	SC	SC	2 copies	SC	Chimpanzee, rat, pig SC	Skaletski et al. 2003; Pearks Wilkerson et al. 2008; Kuroki et al. 2006; Quilter et al. 2002
ZNF33bY	MC	not found	not found	not found	not found	

SC: single copy, MC: multicopy

Heterochromatic transcripts

One of the most intriguing findings of this study is the discovery that three expressed sequences (*ETY3, ETSTY6* and *ZNF33bY)* map to the heterochromatic portion of ECAY. This was confirmed by STS content analysis and by BAC and cDNA FISH. All three ESTs are Y-linked only in horses. Transcripts *ETY3* and *ETSTY6* share no homology with the genomes of other mammals while *ZNF33bY* has an autosomal homolog in horse (ECA1: 66.5 Mb; UCSC, http://genome.ucsc.edu/[4]) and other species. At this stage it is not clear whether the sequences are translated because no ORFs were found (Table 3). This is the first time that transcriptional activity has been found in Y chromosome heterochromatin and has not yet been reported for any other species including humans (Skaletsky et al. 2003).

Taken together, we have constructed a comprehensive sequencing-ready map for the horse Y chromosome and generated a detailed catalogue of 36 ECAY-linked genes. This is a unique resource to further investigate the Y-linked component of male fertility in horses. Identification of the potential ECAY candidate genes for stallion fertility will be the main focus of the next part of this Dissertation.

CHAPTER III

EVALUATION OF THE Y-LINKED GENES AS POTENTIAL CANDIDATES FOR STALLION FERTILITY

INTRODUCTION

General observation from Y chromosomes studies in human, mouse and a few other mammals is that all species share a core set of X-degenerate genes. In addition, they also contain a number of multicopy genes. Some of the multicopy genes are common between species, while others are unique to a particular species or a group of related species (Mazeyrat et al. 1998; Liu et al. 2002; Quilter et al. 2002; Skaletsky et al. 2003; Raudsepp et al. 2004b; Hughes et al. 2005; Kuroki et al. 2006; Murphy et al. 2006). In some instances, a few X-degenerate genes have been amplified and acquired testis-limited expression in one species but have remained single copy and ubiquitously expressed in others (Mazeyrat and Mitchell 1998; Schubert et al. 2000b; Skaletsky et al. 2003). Additionally, in some species Y chromosome has recruited new genes from autosomes or mitochondrial genome (Skaletsky et al. 2003; Murphy et al. 2006; Pearks Wilkerson et al. 2008). The most fascinating and unique feature of the Y chromosome, however, is that despite differences in the evolutionary origin of Y-linked genes, most of these genes tend to carry out similar male reproduction related functions in all species studied so far (Affara and Mitchell 2000; Ellis and Affara 2006).

The human Y chromosome harbors 9 ampliconic genes or gene families, *viz.*, *TSPY, RBMY, VCY, XKRY, CDY, HSFY, PRY, DAZ, BPY2* (Skaletsky et al. 2003). Of these, *VCY*, *RBMY, TSPY, XKRY and HSFY* have functional gametologs on the X chromosome (*VCX, RBMX, TSPYL2, XKRX* and *HSFX1*) whereas *CDY* and *DAZ* have autosomal homologs - *CDYL* and *DAZL,* respectively. *TSPY, RBMY* and *HSFY* are Y-linked, multicopy genes also in several other mammals, like chimpanzee, cat, cattle and pig (Vogel et al. 1997a; Vogel et al. 1997b; Quilter et al. 2002; Tessari et al. 2004; Murphy et al. 2006). Human *TSPY* studies provide evidence that this gene might be involved in early spermatogenesis, immediately prior to the spermatogonia-to-spermatocyte transition, it is also known to be involved in early testicular tumorigenesis (Schnieders et al. 1996; Lau et al. 2009). In contrast, mouse *TSPY* has become single copy and subsequently lost its function whereas *TSPY* in closely related rat is still functional (Mazeyrat et al. 1998). Notably, Y-linked homologs for human *DAZ, BPY2*, *CDY* and *PRY* have not yet been found in any of the non-primate mammalian Y chromosomes. Human *DAZ* belongs to a multicopy gene family while *CDY* and *PRY* each have two and *BPY2* has three nearly identical copies in the palindromic region of HSAY. All four genes are specifically expressed in testis. The function of *PRY* is not yet known, *DAZ* is a critical azoospermia factor, *CDY* protein facilitates the replacement of chromatin histones with protamines, and *BPY2* (*VCY2*) encoded protein interacts with ubiquitin protein ligase E3A and may be involved in male germ cell development and male (in)fertility (Wong et al. 2002; Dada et al. 2004; Kimmins and Sassone-Corsi 2005). Absence of these critical male fertility genes on non-primate mammalian Y

chromosomes suggests that other genes or gene families must have acquired similar functions in other species.

Partial sequencing of chimpanzee (*Pan troglodytes*, PTR) Y chromosome (PTRY) has identified at least 5 ampliconic structures and orthologs for 16 human X-degenerate genes (Hughes et al. 2005; Kuroki et al. 2006). It is necessary to emphasize that to date there is no experimental data showing expression profiles or possible fertility-related functions of any of chimpanzee Y-linked genes. However, comparative sequencing analysis of reading frames and splice sites of the 16 human X-degenerate genes in chimpanzee and gorilla indicates that while all genes are intact in gorilla, structural and likely functional changes have taken place in chimpanzee lineage (Kuroki et al. 2006; Perry et al. 2007; Goto et al. 2009). Six PTRY genes, viz., *CYorf15b, TBL1Y, TMSB4Y*, *USP9Y, VCY* and *VCY1B*, have disruptions and mutations causing possible loss of function or production of truncated proteins (Hughes et al. 2005; Kuroki et al. 2006; Perry et al. 2007).

Besides human, detailed functional studies of Y-linked genes have been carried out only in mouse. It is well established that deletions on the long arm of mouse (*Mus musculus*, MMU) Y chromosome (MMUYq) involving highly amplified genes *Sly*, *Asty* and *Orly* cause male infertility (Toure et al. 2005; Ellis et al. 2007; Reynard et al. 2009). Notably, none of these gene families have been found in HSAY or other mammalian Y chromosomes. Analysis of gene expression profiles on a testis transcript microarray recently identified two more critical spermatogenesis-related murine Y-linked genes. Both are located on MMUYp and are significantly upregulated in the testis RNA from

mice with the MMUYq deletions (Ellis et al. 2005; Ferguson et al. 2009). One of the genes - *H2al2y*- encodes a novel histone family protein associated with centromeric heterochromatin during spermatogenesis, and has related genes on the X chromosome (*H2al1*) and an autosome (*H2al2*). Another gene - *AK006152* - is not related to any known genes and appears to be MMUY-specific without any X-linked or autosomal homologs (Ferguson et al. 2009).

Recently, expression profiles of Y-linked genes were studied also in cats (Murphy et al. 2006; Pearks Wilkerson et al. 2008). Like in humans and mice, the cat Y chromosome contains genes which have multiple copies, are expressed specifically in testis and some of these transcripts such as *TETY1, TETY2* and *FLJ36031Ya* are novel and cat or carnivore specific. Furthermore, many multicopy, but not single-copy genes are significantly downregulated in domestic cat x Asian leopard cat sterile hybrids (W. Murphy unpublished observations) emphasizing the correlation between the function of multicopy Y-linked genes and the status of spermatogenesis.

Proceeding from this overall compelling evidence that the male-specific region of the mammalian Y chromosome is evolutionarily predisposed to accumulate genes important for spermatogenesis, the aim of this study is to investigate whether Y-chromosome genes are associated with fertility also in stallions. We have identified 36 genes and ESTs in the horse Y chromosome (ECAY). All genes will be analyzed for their expression profiles in a panel of different body tissues including testis. The genes/ESTs showing testis-specific expression will be used for qRT-PCR to compare their expression in the testes of normal fertile stallions and stallions with various fertility

problems to identify potential candidate genes for stallion fertility. These goals will be achieved through the following specific objectives.

OBJECTIVES

1. Analyze the expression profiles of all ECAY genes/ESTs in a panel of normal, adult body tissues to identify candidate genes for stallion fertility.
2. Compare the expression levels of all testis-specific genes/ESTs between fertile and infertile/subfertile stallions to identify genes that are differentially expressed between the two groups of animals.
3. Generate full-length or nearly full length cDNA sequences for testis-specific ECAY genes/ESTs to acquire knowledge about their structure.

This research will identify ECAY genes that are differentially expressed in infertile/subfertile stallions compared to normal animals and can hence, be considered as candidate genes for stallion fertility. The results are expected to establish an important foundation for the development of diagnostic tests for reproductive disorders in stallions in the future.

MATERIALS AND METHODS

Collection of tissue samples from normal horses

Fresh necropsy samples of nine tissues (brain, kidney, heart, skeletal muscle, liver, lungs, spleen, seminal vesicle and testis) were obtained from two reproductively

normal adult male horses. Several pieces (approximately 0.5 cm^3) of each tissue were collected in 1 ml of RNA-later (Ambion) to avoid degradation of RNA and stored at -80 ^oC until needed. These samples were used for isolation of messenger RNA (mRNA) and total RNA.

Isolation of messenger RNA (mRNA) from horse tissues

Messenger RNA (mRNA) was isolated using Fast Track 2.0 mRNA isolation kit (Invitrogen) following the manufacturer's instruction with minor modifications (described in Chapter II).

Primer design for Reverse Transcriptase PCR (RT-PCR)

Primers for RT-PCR were designed, if possible, from two neighboring exons flanking an intron, using Primer 3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi[11]). In cases where cDNA sequences were short or it was not possible to determine exon/intron boundaries, intra-exonic primers were used. Details about all primers and PCR conditions are presented in Table 5.

Table 5: Information about primers, reaction conditions and expected product sizes for RT-PCR analysis of all ECAY genes.

Gene Symbol	RT-PCR primers 5′ - 3′	Same as genomic primers	Ta, °C	cDNA product size, bp	Genomic DNA product size, bp
ACTB	F: CCCAGATCATGTTTGAGACCT R: CCTCGTAGATGGGCACAGT	No	58	144	585
AMELY	F: CCAACCCAACACCACCAGCCAAACCTC CCT R: AGCATAGGGGGCAAGGGCTGCAAGGG GAAT	Yes	65	160	160
CUL4BY	F:TGTGGGGTTCGTGTGAAATA R:CAAGGATCGCTGGGTCTTAC	Yes	58	172	172
CYorf15	F:CTAGGTGGCGACGCAAGTGA R:TGCACTCCGATTCTTGTTGA	No	58	358	2342
DDX3Y	F:CTCGAGATCCAAAACTGCTG R:GCTGGTCTGGACCTGAACTC	No	58	181	4378
EIF1AY	F:GATCGTGGCCTTCTGACATT R:TTATTTTTGGGCATGGTGGT	Yes	58	187	187
EIF2s3Y	F:GAGCCATCTGTGTGATCGTC R:TATTCCTGGCCCTAAGCACA	Yes	58	223	223
EIF3CY	F:CCCAAGCAGGGTACCTATGG R:GGACAGAAGTGACGCAATCA	Yes	58	134	134, 230
ETSTY1	F:GACGGACGACCTTGTGTTTT R:CTAGTGGCGAGTCCTTTTGG	Yes	58	234	234
ETSTY2	F:ATCATCGTGGAAAGCCTCAC R:AGTGCTGAAGAGGCTGTGGT	Yes	58	223	223
ETSTY3	F:TTACATTTGTTGCGCCATGT R:GCCCAAAGAAGTAACCGACA	Yes	58	134	134
ETSTY4	F:TAAGGCTTCCCTCCTCCAAT R:CCAGTGACCCGACATACTGA	Yes	58	175	175
ETSTY5	F:CAAAACCAAGAGGAGGACCA R:CTCCAGAGGCAGGTACTTCG	Yes	58	210	210

Table 5 continued

Gene Symbol	RT-PCR primers 5′ to 3′	Same as genomic primers	Ta, °C	cDNA product size, bp	Genomic DNA product size, bp
ETSTY6	F:ACATGGCGCAACAAATGTAA R:TAGCTGTTTGCTGCAGTGCT	Yes	58	245	245
ETY1	F:TCCAGAGCAACAACAGCAAC R:CATCAGTCTGCCCAAACCTT	Yes	58	127	127
ETY2	F:TAAGGCTTCCCTCCTCCAAT R:CCAGTGACCCGACATACTGA	Yes	58	150	850
ETY3	F:TTTTGGCTTGTGTCTTTCTCTG R:ATAGGGCCAGACTTTCACAGC	Yes	58	350	350
ETY4	F:TGGGGATATTGGCTTAGCTG R:CTGGGAGCACGTCTGTATCA	Yes	58	180	180
KAL1Y	F:AGGCACAGTCTTAGGGCAAA R: TTTTGGCATTCCCTTCTCTG	Yes	58	231	231
KDM5D (SMCY)	F:AACAGCGAGCCAATGTTTTT R:GCAAAATTCTGGGAAATCCA	Yes	58	191	400
MAP3K7IP3Y	F: GTGGAATCCCTATTGCTAAAGTTAC R: CCAGAGAGCTGTGACCAAG	Yes	58	138	138
MT-ND1Y	F:CCCTCCGCTTTCCTAGACC R:CAACGATGGCTTGAAAGGAT	Yes	58	100	100
NLGN4Y	F:GGGGATCCATCTTTGTGTTG R:GTCACACAGCAGGCTCTGAC	Yes	58	156	156
RBMY	F:TTCGGCCTTCTCTTTCACAT R:ACTCAAGCAGCCGAAATGAT	Yes	58	180	180
RFX5Y	F:ACCCTTAGGGGGAAAAATCC R:TTTCGTCCCTCAAGTTCCTG	Yes	58	201	201
RPS3AY	F:CCGGAAGAAGATGATGGAAA R:CAAACTTGGGCTTCTTCAGC	Yes	58	179	179, 790
SRY	F:TGCATTCATGGTGTGGTCTC R:ATGGCAATTTTTCGGCTTC	Yes	58	200	200

Table 5 continued

Gene Symbol	RT-PCR primers 5′ to 3′	Same as genomic primers	Ta, °C	cDNA product size, bp	Genomic DNA product size, bp
STS-Y	F:TGTGTGTTTCTGTCATGGGGATTACATC R:CAGACAATGTTTCCCAGTGACAATTGATTA	n/a	58	n/a	210 (M/F)
TBL1Y	F: CACTCGAAACCAATGGAA R: TTCCATATCCTGGCAGTCGA	n/a	50	n/a	500 (M/F)
TMSB4Y	F:ACCCACCCAGCCTCTTACTT R:TTGAAGAAGACGGAAACGC	No	58	334	1123
TSPY	F:GAAGTCAGGCACACCAGTGA R:TAAGGCTGCAGTTGTCATGC	Yes	58	189	280
UBE1Y	F:TGGCCAACTCACGGCTGATCCAA R:CTTCTCCACTCACCCTACTTGGG	Yes	58	210	210
USP9Y	F:GGTTATGAAATGGTCTCTGC R:CGAGTCTGTCCATCAGGAGTC	Yes	58	228	228
UTY	F: CAGCTGTTTTCGGTGATGAG R: GCCTCCTTCTCTTCGGTTG	Yes	54	110	110
YIR2	F:AGGGTTGGGCTAAGTCACCT R:ACCTTGGATCCAGACTCACG	Yes	58	170	170
ZFY	F:TGAGCTATGCTGACAAAAGGTG R:TCTTTCCCTTGTCTTGCTTGA	Yes	58	186	186
ZNF33bY	F:CCACAGCAAATACAGGAGCA R:GTCTGACTCCTCCCCCTTTC	Yes	58	242	800, 3000

Reverse Transcriptase PCR for expression analysis

RT-PCR was carried out using Superscript III One-Step RT-PCR System and Platinum Taq DNA polymerase (Invitrogen) in 15µl reactions containing 40 pmol of each primer and 10 ng of DNase-treated mRNA. RT-PCR with STS primers from a non-transcribed sequence served as a control for mRNA possible contamination with genomic DNA. A housekeeping gene, beta actin (*ACTB*) was used as a control for RT-PCR reactions. RT-PCR started with 30-min incubation at 50ºC, followed by 2-min hot-start incubation at 94ºC and proceeded with 30 cycles as follows: 15 sec 94ºC, 30 sec 58ºC, 1 min 68ºC and final extension 5 min at 68ºC. Genomic controls were run simultaneously with mRNA samples and RT-PCR products were visualized on 2.0% agarose gels.

Isolation of total RNA from horse testis

Testis tissue samples were collected from normal and infertile/subfertile horses (Table 9) and stored in RNA-later (Ambion) at -80ºC. Total RNA was extracted from testis tissues using RNeasy Mini Kit (Qiagen) following manufacturer's protocol. In brief, tissue samples were carefully transferred from RNA-later into 600 µl of RLT buffer (with β-Mercaptoethanol) and homogenized using a glass homogenizer. The homogenate was centrifuged at 10,000 x g for 3 min and the supernatant was transferred carefully into a new tube. An equal volume of 70% ethanol was added to the clear lysate and mixed immediately. The solution was transferred to RNeasy spin column and centrifuged at 10,000 x g for 1 min to bind the RNA to the column. After on-column

DNA digestion with DNase in RDD buffer at room temperature for 15 min, 700 µl of RW1 buffer was added to the column and centrifuged at 10,000 x g for 1 min. The columns were washed twice with RPE buffer. RNeasy spin columns were then placed into fresh collection tubes, 50µl of RNase-free water was added directly into column and left for 2 min. Finally, the column was centrifuged at 10,000 x g for 1 min to elute total RNA. The quality and quantity of total RNA was analyzed using NanoDrop spectrophotometer and Agilent Bioanalyzer (Agilent, CA) with the RNA 6000 Nano chip kit (Agilent, CA).

Primer design for quantitative Real Time PCR (qRT-PCR)

Primers for qRT-PCR were designed, if possible, from two neighboring exons flanking an intron or from one exon using Primer 3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi[11]). The product size of primers varied from 100 to 250 bp (Table 6) which is the ideal product range for qRT-PCR. The primers were first optimized in 10 µl volume PCR reactions containing 1X buffer (Sigma Aldrich), 0.3 pmol of each primer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$ and 0.25 units JumpStart REDTaq DNA polymerase (Sigma Aldrich), and products were resolved on 2.0% agarose gels. Only the primers that amplified single products of the expected size and gave no extra bands or primer-dimers were used for qRT-PCR.

Table 6: Information about primers, reaction conditions and expected product sizes for qRT-PCR analysis of multicopy and testis specific ECAY genes.

Gene Symbol	qRT-PCR primers 5′ to 3′	Same as RT-PCR primers	Ta, °C	qRT-PCR product size, bp
ACTB Housekeeping control	F: CCCAGATCATGTTTGAGACCT R: CCTCGTAGATGGGCACAGT	Yes	58	144
DDX3Y single copy control	F:CTCGAGATCCAAAACTGCTG R:GCTGGTCTGGACCTGAACTC	Yes	58	181
ETSTY1	F:GACGGACGACCTTGTGTTTT R:CGCTCACAGATGACAGTAGCA	No	58	165
ETSTY2	F:AACCAGGAAGCCCAGTTACA R:GTTTGCCTCTTTGGATGAGC	No	58	227
ETSTY3	F:CCTAACCGACAGCCAAAGAG R:GCCCAAAGAAGTAACCGACA	No	58	198
ETSTY4	F:TAAGGCTTCCCTCCTCCAAT R:CCAGTGACCCGACATACTGA	Yes	58	175
ETSTY5	F:CGAGGTCAAAACCAAGAGGA R:CTCCAGAGGCAGGTACTTCG	No	58	216
ETSTY6	F:GCAGTAGGCAGTCGAAGGAA R:AATGGAAGAAGGGGCACAAT	No	58	153
GAPDH Housekeeping control	F: CCTTCTCTTGCTGGGTGATTG R: GACAATGAATTTGGCTACAGCA	Not used in RT	58	103
RBMY	F:TTCGGCCTTCTCTTTCACAT R:ACTCAAGCAGCCGAAATGAT	Yes	58	180
TSPY	F:GAAGTCAGGCACACCAGTGA R:TAAGGCTGCAGTTGTCATGC	Yes	58	189
UBE1Y	F:TGGCCAACTCACGGCTGATCCAA R:CTTCTCCACTCACCCTACTTGGG	Yes	58	210

Gene expression analysis using qRT-PCR

The LightCycler® 480 DNA SYBR Green chemistry technique (Roche Applied Sciences, (Gibson et al. 1996; Heid et al. 1996) was utilized to analyze the expression profiles of testis-specific ECAY genes/ESTs in testis of normal and subfertile/infertile horses. Total RNA was directly reverse transcribed to cDNA using the TaqMan Reverse Transcriptase reagents (Applied Biosystems). Serial dilutions of pooled cDNA samples were used to generate relative standard curves and test the amplification efficiency of each primer. In addition to testis-specific genes of interest, one Y-linked X-degenerate gene (*DDX3Y*) and two autosomal house-keeping genes Beta Actin (*ACTB*) and Glyceraldehyde Phosphate Dehydrogenase (*GAPDH*) were used as references for qRT-PCR. For each qRT-PCR assay a 20µl reaction containing ~100 ng of cDNA, 1x LightCycler® 480 DNA SYBR Green Master I Mix (Roche Applied Sciences) and 10µM primers was amplified on a LightCycler® 480 Detection System (Roche Applied Sciences).

qRT-PCR data analysis

qRT-PCR raw data was initially analyzed using LightCycler® 480 v.1.2 software to generate standard curve, evaluate the efficiency of each primer pair and obtain maximum crossing point (Cp) values for all experimental samples. The software analysis of real-time PCR data is generally based on a method called "cycle-threshold" method. The cycle-threshold is defined as the fractional cycle number in the log-linear region of PCR amplification in which the reaction reaches fixed amounts of amplicon DNA. One

method for calculating the cycle threshold value is second derivative method which calculates the fractional cycle where the second derivative of the real-time fluorescence intensity curve reaches the Cp value (Luu-The et al. 2005; Guescini et al. 2008). The data obtained from LightCycler® 480 v.1.2 software were imported to qBase software and analyzed for relative quantitative expression of genes (Hellemans et al. 2007). p-values to test the statistical significance of relative expression differences (fold induction) for each primer set were calculated using SPSS v.17 software at a cut-off threshold of $p<0.05$.

Copy number analysis using qRT-PCR

The primers designed for gene expression analysis by qRT-PCR (Table 6) were used also to analyze the genomic copy number of multicopy genes on a panel of 8 normal and 12 infertile/subfertile stallions. First, primers for each gene were used to amplify from DNA of a normal fertile stallion. PCR products were directly cloned into TOPO-TA cloning vector and transformed into TOP-10 chemically competent *E. coli* cells (Procedure described in CHAPTER II). Plasmid DNA containing inserts of ECAY genes was isolated using QuickLyse Miniprep Kit (Qiagen) following manufacturer's instructions. Standard curve was generated using serial dilutions of plasmid DNA of each gene as single copy control and efficiency values were calculated. Duplicate samples of 50 ng of genomic DNA from all 20 animals were used as templates for each Real Time copy number experiment. Each experiment was replicated at least twice. The standard curve generated for each transcript was used to calculate the concentration of

the genomic samples. Copy number of each gene per 3.65 pg of DNA (amount of DNA in a haploid cell) was calculated using a copy number calculation formula provided by Roche.

Primer design for RACE

Rapid Amplification of cDNA Ends (RACE) was performed using GeneRacer^{TA} Kit (Invitrogen). For the 5′ end of each partial cDNA sequence, one 5′ reverse primer and one 5′ nested reverse primer were designed using Primer 3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi[11]) in conjunction with GeneRacer 5′ forward primer and GeneRacer 5′ nested forward primer, respectively (provided with the kit) (Table 7). Similarly, for the 3′ end of each partial cDNA sequence, one 3′ forward primer and one 3′ nested forward primer were designed in conjunction with GeneRacer 3′ reverse primer and GeneRacer 3′ nested reverse primer, respectively (provided with the kit). In both cases, manufacturer's instructions were taken into consideration while designing the primers. The nested primers had sequence overlap with the respective gene-specific primers to obtain more specificity of the genes subjected to RACE.

Table 7: Information about RACE primers.

Primers	Sequences 5′ to 3′
ETSTY1-RACE-5′R	TCAGGGCCAAAAACACAAGGTCGTC
ETSTY1-RACE-N-5′R	AAAACACAAGGTCGTCCGTCCAC
ETSTY1-RACE-3′′F	CTTTCCAGATCTGCTCCGTGTGACT
ETSTY1-RACE-N-3′F	GATCTGCTCCGTGTGACTGGTGCTA
ETSTY2-RACE-5′R	GGAGGTGTGGTTACTCTCCTTGGGTTGA
ETSTY2-RACE-N-5′R	TGGTTACTCTCCTTGGGTTGAGTGC
ETSTY2-RACE-3′F	CAGCCCAAAGAAGTAACCGACACG
ETSTY2-RACE-N-3′F	CAGAAAGCCAAACCACAGCCTCTTC
ETSTY3-RACE-5′R	CCCGGGAAGGCTAACCGGAAACTAT
ETSTY3-RACE-N-5′R	CTCTGCCAAAGCTGACTGAGGAAAC
ETSTY3-RACE-3′F	ATGTGTAGGGCCAGTTGAACAGCAG
ETSTY3-RACE-N-3′F	CCAAAGAAAAACCCAGCCTCAGC
ETSTY4-RACE-5′R	GCTGTGGAGGTTATGGTTTGCCCTTG
ETSTY4-RACE-N-5′R	CGGTTTGGTTAGTTCGTGGAGGTT
ETSTY4-RACE-3′F	AACCGTACCTTCTGCAGCAACCAG
ETSTY4-RACE-N-3′F	AGGGAGCCTAGCCACAAACTGCAC
ETSTY5-RACE-5′R	TGCTCCAACTTGCCTCTCTTTTGCAG
ETSTY5-RACE-N-5′R	CTCTGCCTTTACGCATTCCCTCATAC
ETSTY5-RACE-3′F	AAGCTCACGGCATAACGGGCTAGTA
ETSTY5-RACE-N-3′F	CTGGAGGCGACGAAGTACCTCAGAA
ETSTY6-RACE-5′R	GGGTCCTGGGTCAGTTACCACAGAGG
ETSTY6-RACE-N-5′R	GCAGATGTAGTCTGGCTTCCTGGAT
ETSTY6-RACE-3′F	ATACAGACGTGCTCCCAGGCACTT
ETSTY6-RACE-N-3′F	AGGACACTGGTGGCCTTGGTCTCT
TSPY-RACE-5′R	TGGCTGACATCTGGGGGTGGTTCA
TSPY-RACE-N-5′R	TGGCTGACATCTGGGGGTGGTTCA
TSPY-RACE-3′F	GCAGCCGGAGACGGGACTGAGAGTAGG
TSPY-RACE-N-3′F	GCAGCCGGAGACGGGACTGAGAGTAGG
RBMY-RACE-5′R	GGGGGAGAGGCGTATATTGGCTTT
RBMY-RACE-N-5′R	CGATGTGAAAGAGAAGGCCGAAG
RBMY-RACE-3′F	CAGAGGGAGTGCTCGTGGTG
RBMY-RACE-N-3′F	GCTCGTGGTGGTGGCGCACC
GeneRacer™ 5′ Primer F	CGACTGGAGCACGAGGACACTGA
GeneRacer™ 5′ Nested (N)Primer F	GGACACTGACATGGACTGAAGGAGTA
GeneRacer™ 3′ Primer R	GCTGTCAACGATACGCTACGTAACG
GeneRacer™ 3′ Nested (N) Primer R	CGCTACGTAACGGCATGACAGTG

Generation of full length cDNA using RACE

To generate full length cDNA from a partial cDNA sequence, RACE was carried out separately for 3′ and 5′ ends. 3′ RACE was done using 3′ RACE primers and GeneRacer systems (Invitrogen) according to manufacturer's specifications. Approximately, 1 μg of total testis RNA was directly used for reverse transcriptase PCR for 3′ RACE. For 5′ RACE, 1 μg of total RNA was dephosphorylated, decapped and GeneRacer RNA oligo was ligated to the full length mRNA. The dephosphorylation of total RNA used for 5′ RACE helped to eliminate the truncated mRNA and non-mRNA from subsequent ligation with the GeneRacer oligo. However, full-length mRNA remained intact after dephosphorylation due to the presence of 7mG cap at the 5′ end. This 5′ cap was removed by pyrophosphatase keeping the full-length mRNA structure intact. This treatment also left available a 5′ phosphate that is required for the ligation to the GeneRacer RNA oligo (provided with the GeneRacer kit).

After the modification of 5′ mRNA end, both 5′ and 3′ testis RNA were separately subjected to reverse transcription to synthesize cDNA as templete for RACE PCR using random primers for 5′ end and oligo dT primers for 3′ end reactions. The 5′ and 3′ RACE-ready cDNA were then used for RACE PCR amplification using gene-specific primers (Table 7). The first-round PCR cycling conditions for RACE were: hot-start at 94°C for 2 min; 5 cycles of 94°C for 30 sec, 72°C for 1 min; 5 cycles of 94°C for 30 sec, 70°C for 1 min; 20 cycles of 94°C for 30 sec, 65°C for 30 sec and 68°C for 1 min. A final extension of 10 min at 68°C completed the reaction. The PCR product obtained from the first round of amplification was diluted 10 times (1:10 dilution) and 1

µl of the dilution was used as a template for the second round of amplification (Nested RACE PCR) with nested primers. Nested RACE PCR was performed using the cycling conditions as follows: hot-start at 94°C for 2 min; 20 cycles of 94°C for 30 sec, 65°C for 30 sec and 68°C for 2 min followed by a final extension of 10 min at 68°C. Products of nested RACE PCR were visualized on a 2% agarose gel. The bands were cut from gel and PCR products were eluted using S.N.A.P. columns provided with the kit. PCR products were thereafter cloned using TOPO TA Cloning Kit for Sequencing (Invitrogen). Transformed cells were plated on LB agar with 50 µg/ml ampicillin. Colonies were picked after an overnight incubation at 37°C and cultured overnight at 37°C in LB medium with 50 µg/ml ampicillin. Plasmid DNA was extracted using either REAL Prep 96-well Kit (Qiagen), for large number of clones or QuickLyse Miniprep Kit (Qiagen), for a few number of clones depending on the number of clones to be analyzed at a time. Plasmid DNA was digested with EcoR1 and visualized on a 1% agarose gels. Plasmids containing appropriate size inserts were sequenced using universal primers as described in Chapter II.

RESULTS

Gene expression analysis by RT-PCR

Reverse transcriptase PCR using primers for 34 ECAY genes and ESTs was carried out on nine equine body tissues. Primers for *TBL1Y* and *STS-Y* amplified both male and female genomic DNA and were not used for analysis. In total, 31 genes/ESTs

showed expression in one or more tissues while no expression was observed for *ETY3, KAL1Y* and *AMELY*. This explains why the latter two were not found from cDNA selection. Expression profiles could be divided into three distinct categories: i) ten multicopy genes/ESTs (*ETSTY1-6, RBMY, TSPY, UBE1Y, ZNF33bY*) were expressed only in testis (Fig. 8a); ii) six genes/ESTs (*CUL4BY, ETY1, NLGN4Y, RFX5Y, SRY, YIR2*) showed intermediate expression being expressed in testis and in a few other tissues; two genes in this group - *YIR2* and *SRY* - were expressed predominantly in testis (Fig. 8b) and, iii) 15 genes/ESTs (*CYorf15, DDX3Y, EIF1AY, EIF3CY, EIF2s3Y, ETY2*, *ETY4, MAP3K7IP3Y, MT-ND1Y, RPS3AY, KDM5D (alias SMCY), TMSB4Y, USP9Y, UTY, ZFY)* were expressed in all nine tissues (Fig. 8c, Table 8). The 10 multicopy genes/ESTs showing testis-specific expression (Table 8) were considered as potential candidates for stallion fertility and were used for quantitative gene expression studies across a panel of testis RNA isolated from normal and infertile/subfertile stallions.

Table 8: Summary of the expression profiles of ECAY genes and ESTs.

Gene Symbol	Expression Pattern	Expression in horse tissues	Expression in human tissues	Copy number
AMELY	No expression	-	Expressed in tooth enamel	Single copy
CUL4BY	Intermediate	Predominantly testis; limited in kidney, heart, spleen	Not found in human	Multicopy
CYorf15	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Ubiquitous	Single copy
DDX3Y	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Ubiquitous	Single copy
EIF1AY	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Ubiquitous	Single copy
EIF2s3Y	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Not found in human	Single copy
EIF3CY	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	n/a	Single copy
ETSTY1	Testis-specific	Testis	Not found in human	Multicopy
ETSTY2	Testis-specific	Testis	Not found in human	Multicopy
ETSTY3	Testis-specific	Testis	Not found in human	Multicopy
ETSTY4	Testis-specific	Testis	Not found in human	Multicopy
ETSTY5	Testis-specific	Testis	Not found in human	Multicopy
ETSTY6	Testis-specific	Testis	Not found in human	Multicopy

Table 8 continued

Gene Symbol	Expression Pattern	Expression in horse tissues	Expression in human tissues	Copy number
ETY1	Intermediate	Predominantly testis and liver, limited in brain, kidney, heart, lungs, spleen	Not found in human	Multicopy
ETY2	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Not found in human	Single copy
ETY3	No expression	-	Not found in human	Multicopy
ETY4	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Not found in human	Multicopy
KAL1Y	No expression	-	Pseudogene, not expressed	Single copy
KDM5D (SMCY)	Ubiquitous	Ubiquitous	Ubiquitous	Single copy
MAP3K7IP3Y	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Not found in human	Single copy
MT-ND1Y	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Pseudogene, not expressed	Single copy
NLGN4Y	Intermediate	Testis, seminal vesicle, brain	Testis, prostate, brain	Single copy
RBMY	Testis-specific	Testis	Testis	Multicopy
RFX5Y	Intermediate	Testis, kidney, liver, spleen	Expression data not available	Single copy
RPS3AY	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Expression data not available	Single copy
SRY	Intermediate	Predominantly testis; limited in kidney, seminal vesicle	Predominantly testis	Single copy
STS-Y	n/a	n/a	Not found in human	Single copy
TBL1Y	n/a	n/a	Ubiquitous	Single copy
TMSB4Y	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Ubiquitous	Single copy

Table 8 continued

Gene Symbol	Expression Pattern	Expression in horse tissues	Expression in human tissues	Copy number
TSPY	Testis-specific	Testis	Testis	Multicopy
UBE1Y	Testis-specific	Testis	Not found in human	Multicopy
USP9Y	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Ubiquitous	Single copy
UTY	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Ubiquitous	Single copy
YIR2	Intermediate	Testis, heart; limited in kidney, liver, lungs, spleen	n/a	Multicopy
ZFY	Ubiquitous	Brain, kidney, heart, skeletal muscle, liver, lung, seminal vesicle, testis	Ubiquitous	Single copy
ZNF33bY	Testis-specific	Testis	Expression data not available	Multicopy

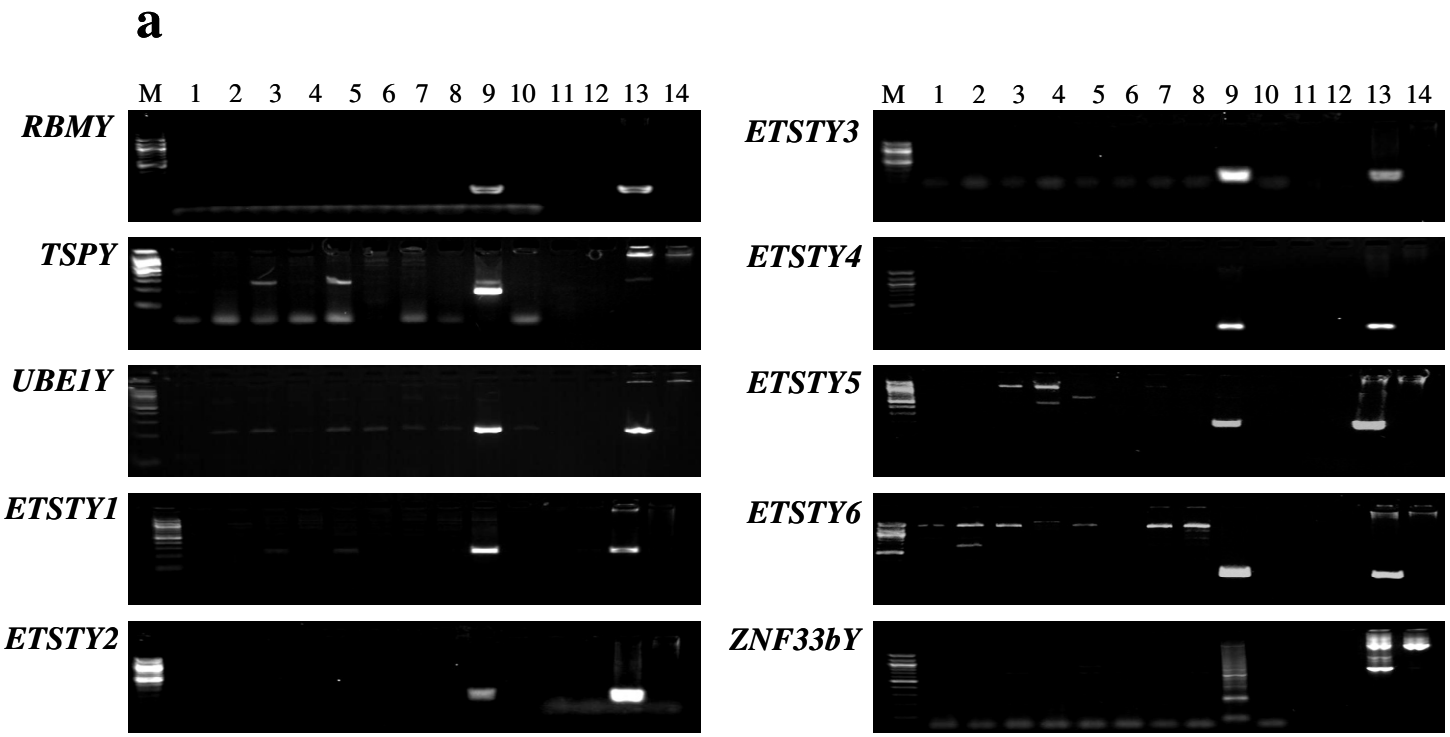

Figure 8: RT-PCR results showing the expression of ECAY genes in nine body tissues. Lanes 1-brain, 2-kidney, 3-heart, 4-muscle, 5-liver, 6-lung, 7-spleen, 8-seminal vesicle, 9-testis, 10-no mRNA control, 11-no RT control, 12-no genomic DNA control, 13-male genomic DNA control, 14- female genomic DNA control, M-100 bp marker. **a.** Testis-specific expression. M: molecular markers.

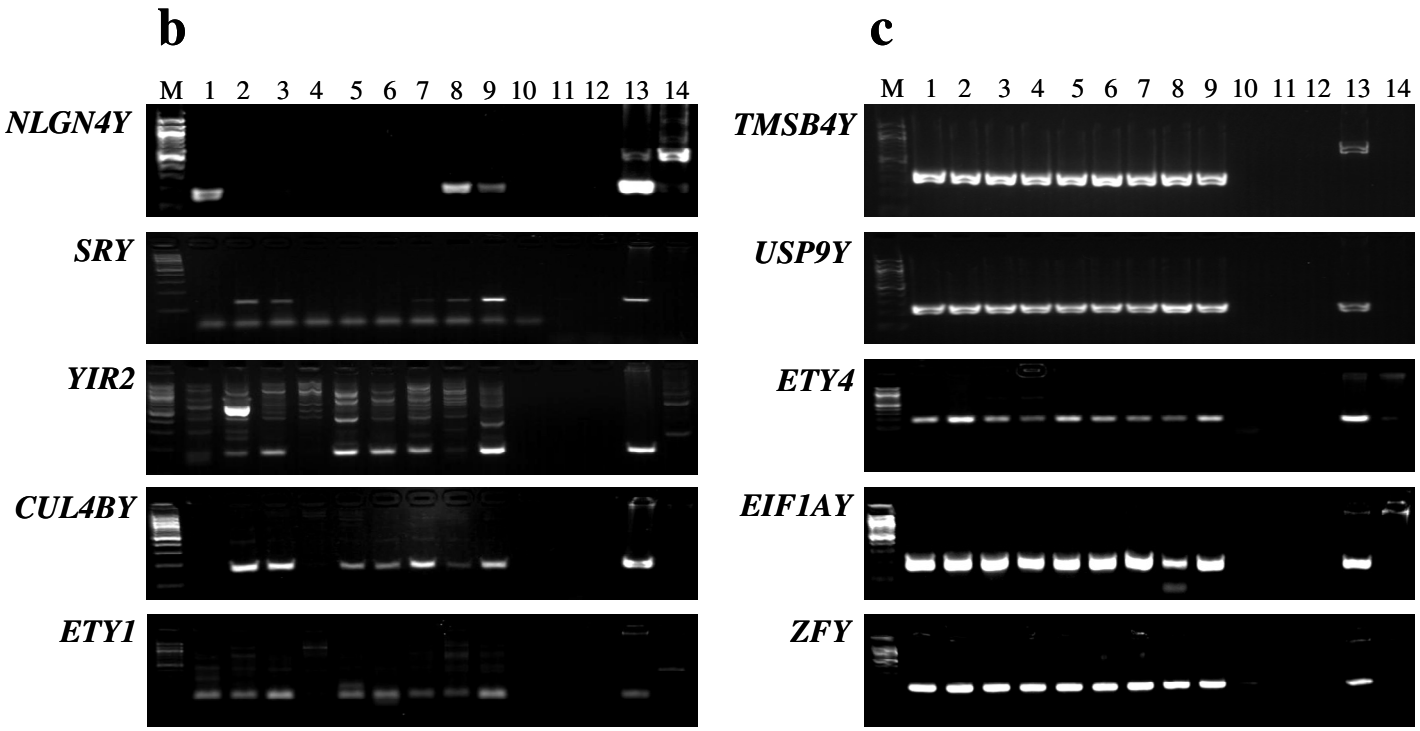

Figure 8 (contd.)
b. Intermediate expression; **c.** Ubiquitous expression.

Comparison of gene expression between fertile and subfertile/infertile stallions

According to our hypothesis, transcripts with testis-limited expression and with multiple copies on the horse Y chromosome are potential candidates for stallion fertility. Therefore, all 10 multicopy and testis-specific ESTs, *viz*., *ETSTY1-6, RBMY, TSPY, UBE1Y* and *ZNF33bY* were selected for quantitative gene expression studies between normal and subfertile/infertile stallions. *DDX3Y*, a ubiquitously expressed single copy gene, and two autosomal reference genes, *ACTB* and *GAPDH*, were used as controls. All primers were first optimized by regular PCR on a genomic DNA template. All markers, except *ZNF33bY*, amplified one single PCR product of the expected size. Primers for *ZNF33bY* amplified multiple products and designing new set of primers for this gene did not change this pattern. Since the presence of a specific PCR product is extremely important for quantitative real time PCR using SYBR Green Chemistry, *ZNF33bY* was excluded from further study. Testis samples from 10 normal, fertile, unrelated stallions and 14 infertile/subfertile stallions were used for this study (Table 9). Total testis RNA from three normal stallions was pooled and reverse transcribed into cDNA. Sequential 1:2 serial dilutions were used to obtain a range of template cDNA quantities (from 200ng to 6.25ng) for qRT-PCR reactions to generate standard curves for each primer set. The standard curves produced amplification efficiency values for each primer set and were used for final calculations (discussed later). Optimal quantity of template cDNA was about 100 ng, thus, this amount of cDNA from each animal was used in duplicate reactions for qRT-PCR assays.

Table 9: Information about the samples used for gene expression analysis by qRT-PCR.

Sample Id	Category	Breed	Age	Clinical description
Normal 1	Normal, fertile	QuarterHorse	4 yr	Normal
Normal 2	Normal, fertile	QuarterHorse	4 yr	Normal
Normal 3	Normal, fertile	QuarterHorse	4 yr	Normal
Normal 4	Normal, fertile	Thoroughbred	4 yr	Normal
Normal 5	Normal, fertile	Gypsy Vanner	n/a	Normal
Normal 6	Normal, fertile	n/a	n/a	Normal
Normal 7	Normal, fertile	n/a	n/a	Normal
Normal 8	Normal, fertile	QuarterHorse	4 yr	Normal
Normal 9	Normal, fertile	QuarterHorse	2 yr	Normal
Normal 10	Normal, fertile	n/a	2 yr	Normal
Abnormal 1	Infertile/subfertile	Arabian	n/a	Sterile. Normal sperm count but 97% of sperm morphologically abnormal (head abnormalities). Low motility.
Abnormal 2	Infertile/subfertile	QuarterHorse	23 yr	Sterile. Very small testes and very low sperm count. Low percentage of normal sperm.
Abnormal 3	Infertile/subfertile	Standardbred	3 yr	Autosomal trisomy (65, XY+27). Sterile, azoospermia, bilateral cryptorchid.
Abnormal 4	Infertile/subfertile	Connemara	7 yr	Sterile, azoospermia.
Abnormal 5	Infertile/subfertile	n/a	n/a	Bilateral cryptorchid. Both testes abdominal

Table 9 continued

Sample Id	Category	Breed	Age	Clinical description
Abnormal 6	Infertile/subfertile	n/a	n/a	Unilateral cryptorchid
Abnormal 7	Infertile/subfertile	n/a	n/a	Unilateral cryptorchid
Abnormal 8	Infertile/subfertile	n/a	n/a	Unilateral cryptorchid
Abnormal 9	Infertile/subfertile	n/a	n/a	Unilateral cryptorchid
Abnormal 10	Infertile/subfertile	n/a	n/a	Bilateral cryptorchid.
Abnormal 11	Infertile/subfertile	Appaloosa	2 yr	Bilateral cryptorchid. Both testicles deeply abdominal, close to the kidneys.
Abnormal 12	Infertile/subfertile	Mustang	6 yr	Unilateral cryptorchid.
Abnormal 13	Infertile/subfertile	QuarterHorse	1 yr	Unilateral cryptorchid.
Abnormal 14	Infertile/subfertile	Thoroughbred	19 yr	Infertile due to impaired acrosomal exocytosis.

qRT-PCR assays with each set of primers were carried out on a panel of testis cDNA from 24 animals. All assays were repeated at least twice to verify the consistency of results. The cDNA from normal stallion number 2 (Normal 2, Table 9) was used as inter-run calibrator to eliminate plate-to-plate variation in cases where all samples could not be run in a single plate. After each experiment, Cp values for each sample were obtained using 2nd derivative max option within absolute quantification program of Light Cycler 480 software v. 1.2. After importing these data to the qBase software, relative quantification was done using *ACTB* and *GAPDH* as reference genes and the remaining 10 genes as genes of interest. Efficiency values that were calculated for each gene earlier were taken into consideration while analyzing data in qBase. The results demonstrated that the reference genes, *ACTB* and *GAPDH,* and single copy *DDX3Y* were expressed at the same level in all animals (Fig. A2a, b), whereas all multicopy testis-specific genes showed differential expression between fertile and infertile stallions (Fig. A2c, d). In summary, *ETSTY1, ETSTY5, TSPY* and *UBE1Y* were upregulated in infertile individuals while all other genes demonstrated downregulation. This result was further analyzed for statistical significance using SPSS v. 17 software. *p*-values were calculated with Mann Whitney non- parametric test which is specifically designed for sample numbers less than 15 in each group (10 fertile and 14 infertile individuals). Five ESTs *viz*., *ETSTY3-6* and *RBMY* showed statistically significant differential expression between fertile and infertile groups with cut-off *p*-value 0.05 (95% level of confidence) (Fig. 9).

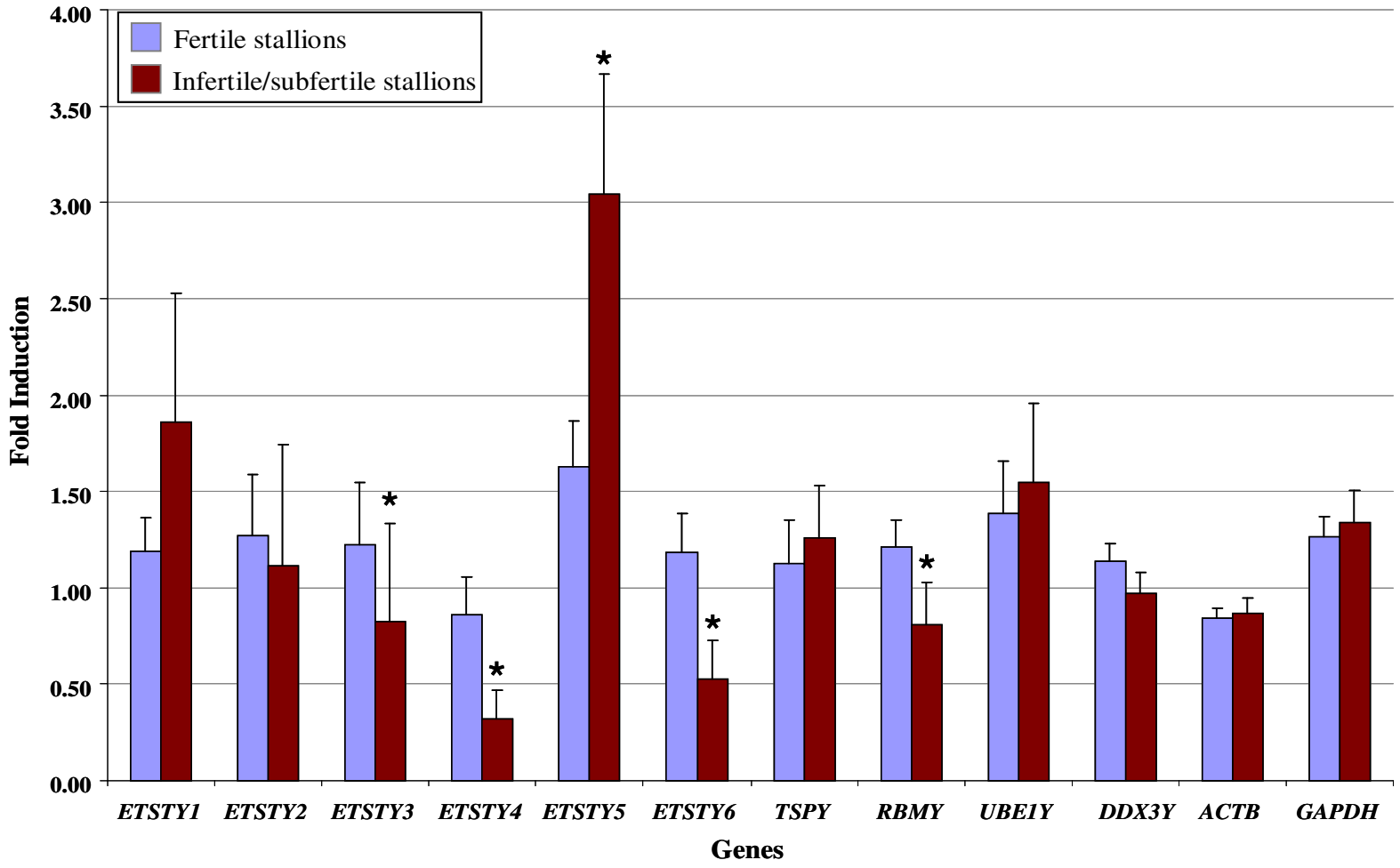

Figure 9: qRT-PCR results showing differential expression of ECAY genes between normal fertile stallions and stallions with fertility problems; * p value < 0.05, statistically significant (Mann Whitney non-parametric test).

Notably, *ETSTY5* was significantly upregulated in infertile individuals while *ETSTY3, 4, 6* and *RBMY* were significantly downregulated. Since the 14 "abnormal" samples were collected from stallions with a broad range of different infertility/subfertility phenotypes there was variation in the expression levels of the same genes across these individuals. For example, *ETSTY1* was downregulated in seven and upregulated in the remaining seven infertile stallions. Interestingly, out of the seven stallions where *ETSTY1* was downregulated, six animals were cryptorchids whereas in three other cryptorchids *ETSTY1* was slightly upregulated (Fig. A2c).

Copy number analysis by qRT-PCR

Copy numbers for nine testis-specific multicopy genes/ESTs (*ETSTY1-6*, *TSPY*, *RBMY*, *UBE1Y*) were analyzed using Real Time PCR absolute quantification method and a panel of genomic DNA from 8 normal stallions and 12 stallions with fertility problems. After obtaining individual copy numbers of each animal, the average copy numbers of each gene across fertile and infertile/subfertile stallions (Table 10) was compared. The average copy numbers of the nine multicopy ECAY genes in fertile and infertile/subfertile stallions are summarized in Table 10 and Fig. 10. Overall, average copy numbers of all studied genes showed a tendency to decrease in infertile animals. For four ESTs, *viz*., *ETSTY2*, *ETSTY6*, *RBMY* and *UBE1Y* this decrease was statistically significant (SPSS v.17, MannWhitney non parametric test for unrelated samples). Genes with the highest copy numbers in both normal and abnormal animals were *RBMY* and *UBE1Y* (Table 10). It is, however, intriguing that during cDNA selection procedure only

one transcript was isolated for *RBMY*, while 54 cDNA clones contained *UBE1Y* (discussed in Chapter II).

Table 10: Results of copy number analysis of multicopy ECAY genes/ESTs.

Gene	Average copy number		Copy number change in subfertile/infertile stallions
	Normal fertile stallions	Subfertile/ infertile stallions	
ETSTY1	29	24	decrease
ETSTY2	6	2	decrease *
ETSTY3	22	14	decrease
ETSTY4	39	33	decrease
ETSTY5	2	1	decrease
ETSTY6	2	1	decrease*
TSPY	49	32	decrease
RBMY	393	155	decrease*
UBE1Y	321	205	decrease*

* statistically significant change (SPSS v.17, MannWhitney non parametric test for unrelated samples)

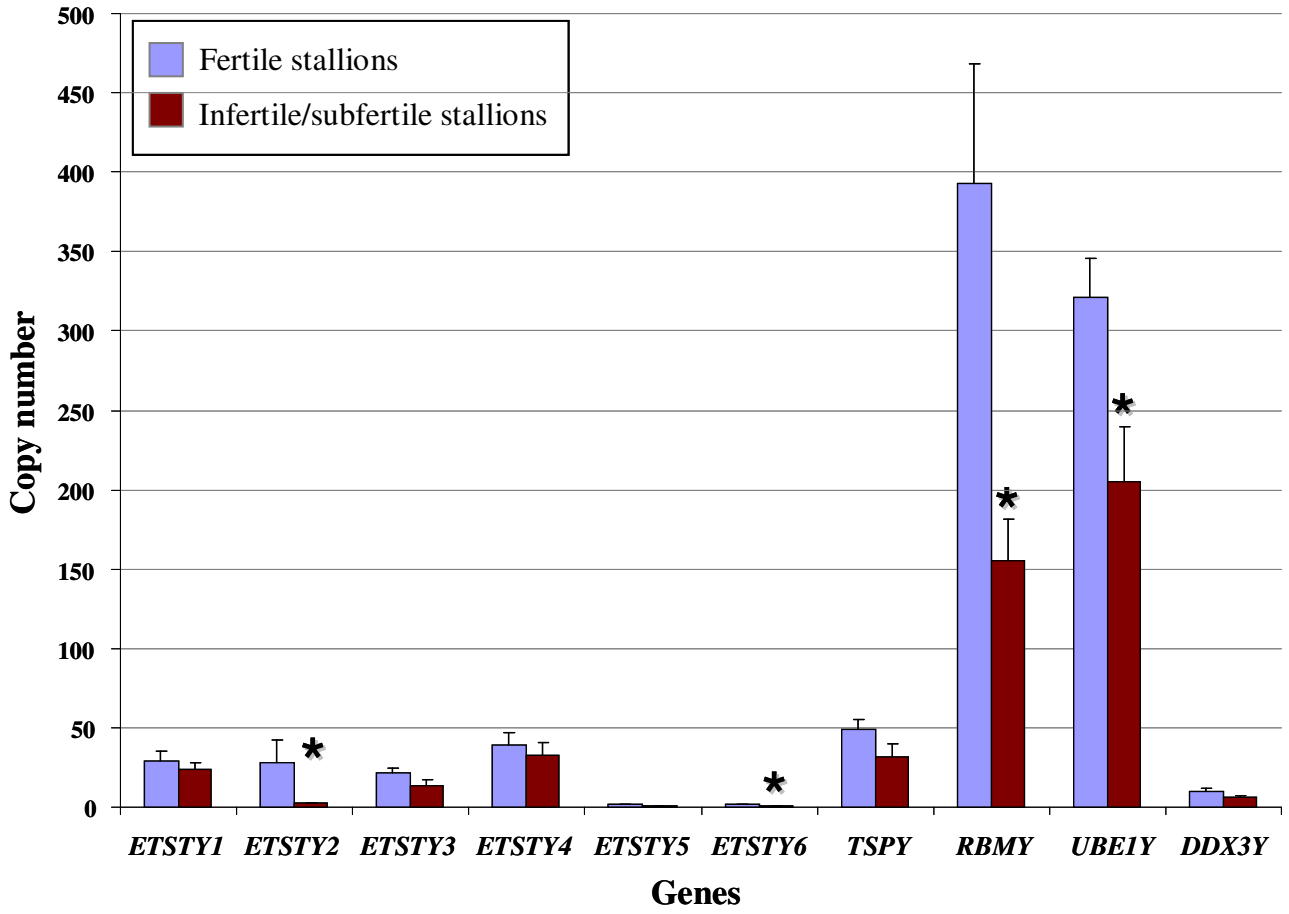

Figure 10: Summary of copy number variation of ECAY genes between normal stallions and stallions with fertility problems; * p value < 0.05, statistically significant (Mann Whitney non-parametric test).

Generation of full length cDNA

Full length cDNA sequences were generated for three testis-specific transcripts - *TSPY*, *ETSTY2* and *ETSTY5* using 5′ and 3′ RACE PCR. The available partial cDNA sequences were used to generate sequences for the 5′ and the 3′ ends. The cDNA RACE PCR products for both ends were gel purified, cloned into plasmid and sequenced. The sequences were quality trimmed, aligned with the original cDNA sequence using Sequencher v1.7. The full length cDNA sequences obtained for the three genes were as follows: *TSPY* 1079 bp, *ETSTY2* is 2323 bp and *ETSTY5* is 1635 bp. The size of the horse *TSPY* cDNA is similar to that of human *TSPY1* which is 1159 bp and the two sequences show 79% identity with an e-value of 6e-97.

Y-specific RACE primers could not be designed for horse *UBE1Y* because its sequence is highly similar to *UBE1X* on the X chromosome. Due to the uncertainty of RACE-PCR procedure, full length cDNAs could not be obtained for the remaining testis specific transcripts. However, for some transcripts the lengths of partial cDNA sequences were increased as follows: *ETSTY1* 448 bp, *ETSTY3* 2591 bp, *ETSTY4* 2452 bp, *ETSTY6* 1066 bp, *RBMY* 712 bp, *UBE1Y* 2924 bp and *ZNF33bY* 3602 bp.

Analysis of novel equine Y sequences

Sequences of the ten novel equine Y-specific multicopy ESTs were aligned with each other using Sequencher V1.7 and assembly parameters 20 for minimum overlap and 90% for minimum match. None of the ten ESTs showed any overlap with each other. However, when assembly parameters were relaxed to 10 for minimum overlap and

75% for minimum match, *ETSTY2, ETSTY3* and *ETSTY6* demonstrated substantial alignment identity with each other. Similarly, when the same three sequences were aligned in pairs using BLAST tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi[9]), short stretches of sequences showed 75-90% identity. The remaining three sequences, however, did not align even at the least stringent assembly parameters and are unique to each transcript.

Presence of open reading frames (ORFs) and evaluation of protein coding ability

The presence of open reading frames was analyzed for all three ESTs for which full length cDNA sequences were obtained. *TSPY* cDNA has eight ORFs of which the longest is 807 nucleotides. The remaining seven ORFs are present at different overlapping regions of the longest ORF and are 693 bp, 375 bp, 366 bp, 342 bp, 330 bp, 276 bp and 237 bp, respectively. All seven ORFs, except the shortest, encode *TSPY* protein of 269, 231, 125, 122, 114, 110 and 92 amino acids, respectively. BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi[9]) analysis showed that horse *TSPY* protein is highly similar to *TSPY* proteins in human and several other mammals. However, the 79 amino acid protein encoded by the shortest equine ORF did not show any significant similarity to *TSPY* proteins in other species. Using NCBI conserved domain database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi[17]) a conserved domain, NAP (nucleosome assembly protein) was identified in horse *TSPY* protein. This domain is conserved in different mammalian *TSPY* proteins, including human, and is important for a diverse spectrum of cellular and molecular functions.

Two non-overlapping ORFs (348 bp and 333 bp) were found in *ETSTY2* and three non overlapping ORFs (219 bp, 381 bp and 180 bp) in *ETSTY5. ETSTY2* can potentilly encode two proteins consisting of 116 and 111 amino acids and *ETSTY5* for three proteins with 73, 127 and 60 amino acids, respectively. No conserved domains were identified for *ETSTY2* and *ETSTY5* full length transcripts using NCBI conserved domain database search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi[17]). ORF analysis with other novel partial cDNA sequences revealed the presence of non-overlapping ORFs in *ETSTY3* (255 bp and 159 bp), *ETSTY4* (237 bp), *ETSTY6* (153 bp, 165 bp and 159 bp), *ETY1* (288 bp and 276 bp) and *ETY4* (123 bp and 129 bp). Each ORF can potentially code for proteins consisting of 85, 53 (*ETSTY3*), 79 (*ETSTY4*), 51, 55, 53 (*ETSTY6*), 96, 92 (*ETY1*), 41 and 43 (*ETY4*) amino acids, respectively. None of these amino acid sequences showed homology with existing non-redundant protein sequences and reference sequence databases using NCBI protein BLAST (BLASTP) algorithm (http://blast.ncbi.nlm.nih.gov/Blast.cgi[9]). This supports our assumption that these are horse-specific expressed sequences and have no counterparts in other mammalian species.

DISCUSSION

Expression analysis of equine Y-chromosome genes

Reverse transcriptase PCR identified 10 ECAY transcripts (*ETSTY1-6, RBMY, TSPY, UBE1Y, ZNF33bY*) that are expressed only in testis. All 10 transcripts are present

in multiple copies on equine Y chromosome. This is in agreement with findings in human, mouse and cat where multicopy Y-linked genes also tend to be expressed only in testis (Skaletsky et al. 2003; Murphy et al. 2006; Pearks Wilkerson et al. 2008). Besides these similarities, there are also lineage-specific differences. For example, *RBMY* is a multicopy and testis-specific gene in horse, human and mouse while *TSPY* is a multicopy and testis-specific gene in horse, human and cat but a single copy in mouse (Schubert et al. 2000a; Schubert et al. 2000b; Skaletsky et al. 2003; Murphy et al. 2006). *UBE1Y* is a single copy and intermediately expressed gene in cat (Murphy et al. 2006), whereas the single copy murine *UBE1Y* is expressed exclusively in testis (Mitchell et al. 1991). In contrast, horse *UBE1Y* is present in multiple copies and like its murine ortholog, has testis limited expression (Fig. 8a). It is likely that horse Y-linked ubiquitin-activating enzyme has acquired functions that are restricted to ubiquitin activation and protein turnover (Levy et al. 2000) specific to germ cell proliferation and is, thus, a part of factors regulating stallion fertility. Notably, this gene has been lost from the Y chromosome in human lineage (Lahn et al. 2001; Skaletsky et al. 2003). Finally, *ZNF33b* is an autosomal gene in human and other mammals, including horse, but has acquired multiple copies and testis-limited expression in the horse Y chromosome. However, *ZNF33bY* demonstrated the presence of multiple bands in horse testis (Fig. 8a). This observation suggests that *ZNF33bY* might be a member of a multicopy gene family rather than being a single gene on the horse Y chromosome. The function of this gene is as yet unknown.

Six ECAY genes and ESTs (*CUL4BY, ETY1, NLGN4Y, RFX5Y, SRY, YIR2)* showed intermediate expression pattern. Until now, *CUL4BY* has been found to be Y-linked only in cat and dog where it is highly amplified and has testis-limited expression (Murphy et al. 2006; Pearks Wilkerson et al. 2008). Horse *CUL4BY* is also multicopy but has a broader expression profile (Table 8). It is possible that equine *CUL4BY* is in a transitional stage from ubiquitous towards testis-restricted expression – a transition from its original housekeeping duties to spermatogenesis-related functions. *NLGN4Y* shows intermediate expression pattern in all species studied so far (mouse, human, cat and horse) and is expressed in reproductive tissues (testis, seminal vesicle, prostate) and brain (Skaletsky et al. 2003). *RFX5Y* is an autosomal gene in other mammals and has acquired a Y-linked copy only in horse. *YIR2* is present in human Y chromosome but its expression data is not known. However, *SRY* - the male sex-determining region on Y seems to be the most interesting intermediately expressed gene. It plays the key role in male sex determination and is expected to be functional at early stages of embryonic development in mammals (Wilhelm et al. 2007b). High levels of *SRY* transcription in adult horse testis and to a lesser extent in a few other adult tissues (Table 8) indicates that *SRY* is functional in adult horses as well. It is likely that *SRY* has acquired some new functions in adult horse testis. Given its high expression level in testis compared to other tissues, it is possible that *SRY* is involved in spermatogenesis or some other male fertility related functions.

The functions of the 15 ubiquitously expressed genes, *viz., CYorf15, DDX3Y, EIF1AY, EIF3CY, EIF2s3Y, ETY2*, *ETY4, MAP3K7IP3Y, MT-ND1Y, RPS3AY, KDM5D*

(*alias SMCY*)*, TMSB4Y, USP9Y, UTY* and *ZFY* are not clear but the expression profile suggests that many of them might be involved in housekeeping activities. However, a few of these genes, *viz.*, *DDX3Y* and *USP9Y* are located in human AZFa region which is one of the critical regions for spermatogenesis (Vogt et al. 1996; Skaletsky et al. 2003). Deletions in AZFa cause Sertoli cell only (SCO) syndrome and affect RNA metabolism of human spermatogenesis (Ditton et al. 2004). Decreased transcriptional activity of testicular *DDX3Y* is associated with severe spermatogenic failure and sperm maturation arrest in humans (Lardone et al. 2007b). In contrast, the role of murine *Ddx3y* in spermatogenesis is not so prominent and it is proposed that biological activity of *Ddx3y* might be taken over by its X-linked and autosomal homologs, *Dbx* and *D1Pas1*, respectively (Vong et al. 2006). *USP9Y* is another human AZFa gene and recent studies suggest that it acts as a fine tuner to improve the efficiency of male fertility related function but is not directly involved in spermatogenesis and sperm maturation (Krausz et al. 2006). This is supported by another recent study that reported a complete deletion of *USP9Y* in a normospermic, fertile man (Luddi et al. 2009). However, murine *Usp9y* gene encodes a potentially functional ubiquitin-specific protease possessing a core promoter region that shares several features characteristic to other testis-specific genes and thus might be involved in important male fertility functions (Hall et al. 2003). More specialized functions than housekeeping have been attributed also to *KDM5D (alias SMCY).* In mouse it is expressed in male meiosis from leptotene spermatocytes to spermatids and might be essential for the progression of spermatogenesis (Akimoto et al. 2008). Possible fertility related functions of the remaining ubiquitously expressed genes

are not known yet but it is likely that at least some of them have roles beyond strictly housekeeping activities.

Male-specific BAC clones containing *TBL1Y* and *STS-Y* were initially isolated using radiation hybrid mapping primers for their X-linked gametologs *TBL1X* and *STS-X,* respectively (Raudsepp et al. 2004a). Since the BACs containing the two genes FISH mapped to ECAY only, they were incorporated in the ECAY contig map (Fig. 6). However, because of high degree of sequence similarity between the X- and Y-linked gametologs of *TBL1* and *STS* we have not been able to design specific primers for *TBL1Y* and *STS-Y*, and therefore the two genes were excluded from expression analysis.

Primary candidate genes for stallion fertility

The ultimate goal of these Y-chromosome studies in horses is to identify markers and regions which are most critical for stallion fertility. The collection of genes and ESTs isolated in this study provides an important foundation to initiate systematic search for Y-linked mutations that might lead to male infertility in horses. Valuable clues about likely candidate genes can be found from human and mouse studies where the link between Y chromosome mutations and male infertility is well established (Ellis and Affara 2006; Ellis et al. 2007; Lange et al. 2008). Although a few single-copy Y-linked genes might be involved in spermatogenesis in human and/or mouse, studies indicate that mutations in testis-specific and multicopy transcripts are primarily associated with impaired spermatogenesis and male fertility. Therefore, all multicopy and testis specific ECAY transcripts have been chosen as the primary candidates (Fig. 8a, Table 8) to start

analysis with. These sequences might share functional similarity with human AZFb and AZFc – regions critically involved in human male infertility (Saxena et al. 1996; Saxena et al. 2000; Ferlin et al. 2003; Skaletsky et al. 2003).

Quantitative expression analysis of primary candidate genes

Expression levels of nine testis-specific, multicopy genes, *viz*., *ETSTY1-6*, *TSPY*, *RBMY*, *UBE1Y* were evaluated to identify their potential involvement in stallion fertility. Five ESTs (*ETSTY3-6*, *RBMY*) showed statistically significant differential expression between fertile and infertile/subfertile stallions. *ETSTY5* was upregulated in all infertile/subfertile individuals depicting that this transcript might be related to impaired spermatogenesis in horses. The remaining four transcripts (*ETSTY3-4*, *ETSTY6* and *RBMY*) were downregulated in infertile/subfertile individuals. *RBMY* is known to play an important role in human and mouse spermatogenesis (Elliott et al. 1996; Elliott et al. 1997; Elliott et al. 1998; Kostiner et al. 1998; Affara and Mitchell 2000). Mice deficient in *Rbmy* develop abnormal sperm and are infertile (Szot et al. 2003). Downregulation of *RBMY* in stallions with fertility problems indicates that this gene might have similar functions also in horses. The three equine specific transcripts, *viz*., *ETSTY3-4* and *ETSTY6* share no homology with other mammalian genes/ESTs and there is no comparative information about their possible functions. However, being equine-specific, multicopy, testis-limited, and downregulated in infertile/subfertile stallions, make *ETSTY3*, *ETSTY4* and *ETSTY6* strong candidate genes for stallion fertility.

Four genes and ESTs (*ETSTY1-2*, *TSPY*, *UBE1Y*) did not show statistically significant differential expression between the fertile and infertile/subfertile groups of stallions but showed up-or downregulation in one or two individual animals. For example, *TSPY* was strongly overexpressed in only one infertile stallion but expressed at the same level in the remaining infertile animals and all normal controls. The stallion with elevated *TSPY* expression had a unique infertile phenotype which was not found in any other animals studied. The stallion had normal sperm count but reduced sperm motility, and 97% of his sperm were morphologically abnormal. It is therefore likely, that *TSPY* is involved in the processes that are responsible for sperm structural integrity and motility. It can be anticipated that if more individuals with sperm structural abnormalities could be analyzed, differential expression of *TSPY* might reach statistically significant values. Inclusion of more animals and more diverse infertile phenotypes into analysis might give significant values also to the differential expression profiles of *ETSTY1*, *ETSTY2* and *UBE1Y* and indicate what kind of role they might have in stallion spermatogenesis.

Because of the contrasting role of *DDX3Y* in human and mouse spermatogenesis (discussed above), this gene was chosen as our X-degenerate, ubiquitously expressed Y-linked control for expression analysis. *DDX3Y* did not show any statistically significant differential expression between fertile and infertile/subfertile individuals and at this stage of research this gene is therefore not considered as a potential candidate for stallion fertility. This observation complies with mouse studies where *DDX3Y* transcript does not show any association with spermatogenesis (Vong et al. 2006).

Finally, we observed individual differences in the expression levels of the same gene within the group of normal animals and within the group of stallions with fertility problems. Given that male fertility is governed by thousands of genes, it is likely that there are additional factors that modulate the expression of the Y-linked genes analyzed in this study.

Effect of copy number variation on gene expression levels and male infertility

The expression and copy number data obtained from qRT-PCR analysis allowed to investigate whether differential expression of a gene is correlated with its copy numbers and whether gene expression and copy numbers are correlated with a reproductive phenotype. It appeared that out of the four genes/ESTs (*ETSTY2, ETSTY6, RBMY* and *UBE1Y*) that demonstrated significant decrease in average copy number in infertile/subfertile stallions, three genes (*ETSTY6, ETSTY2* and *RBMY*) were also downregulated in these individuals. Notably, downregulation of *ETSTY6* and *RBMY* was statistically significant (Fig. 9). Therefore, the lower transcriptional level of the three genes is likely caused by decreased average copy number. In contrast, there was negative correlation between *UBE1Y* copy numbers and transcription level – in infertile/subfertile stallions *UBE1Y* has less copies but its transcription is upregulated. It is possible that elevated transcription compensates for the loss of *UBE1Y* copies, or alternatively, loss of copies in stallions with fertility problems is associated with a regulatory mutation which increases the transcriptional level of this gene.

It is worth to mention that the stallion with sperm morphological abnormalities and elevated levels of *TSPY* transcription (abnormal 1 Table 8, discussed in Chapter III) showed decrease in *TSPY* average copy numbers. The mechanism of such negative correlation between *TSPY* transcription and copy numbers and causes of sperm morphological abnormalities remain currently unclear but provide material for future studies.

Full length cDNA sequences: structure in relation to function

Full length cDNA sequences were obtained for three multicopy, testis-specific genes/ESTs: *TSPY* and two novel Y-linked transcripts, *viz., ETSTY2* and *ETSTY5.*

TSPY is an evolutionarily conserved gene on the Y chromosome of all placental mammals, except mouse (Vogel et al. 1997a; Vogel et al. 1997b). The full-length cDNA sequence of equine *TSPY* showed the presence of a NAP domain which is conserved across all known mammalian *TSPY* proteins. Humans have three different polymorphic TSPY proteins and each contains a conserved NAP domain of ~160 amino acids (Lau et al. 2003). NAP domain binds to Beta Cyclins and core histone proteins and plays role in DNA replication, cell cycle regulation, transcription and chromatin remodeling (Nagata et al. 1995; Compagnone et al. 2000; Chai et al. 2001; Zhang et al. 2001; Canela et al. 2003). Mutations or dysregulation in the members of NAP gene family have been associated with various forms of human cancers (Chai et al. 2001). In normal individuals *TSPY* directs the spermatogonial cells to enter meiosis (Schnieders et al. 1996; Lau 1999) and might have an additional mitotic function in the proliferation of embryonic

gonocytes and adult spermatogonia (Honecker et al. 2004). Consequently, the improper proliferation of spermatogonial cells might affect sperm production and thus, male fertility. Results of this study show high degree of structural similarity between the full length cDNA sequences of human and horse *TSPY* genes. Despite the fact that we were not able to detect statistically significant differential expression of *TSPY* between normal and reproductively challenged stallions, it could be expected that *TSPY* plays important role in germ cell proliferation also in horses. Furthermore, overexpression of *TSPY* in one infertile stallion with sperm morphological defects (see above) indicates that the function of *TSPY* might be more complex than recognized today.

Full length cDNA and the corresponding protein sequences of *ETSTY2* and *ETSTY5* are novel and equine specific and share no significant similarity with other known mammalian sequences. Therefore no comparative analysis of the two ESTs could be done.

Novel ECAY sequences

cDNA selection studies identified 10 novel horse-specific Y-linked transcripts. Six of these (*ETSTY1-6*) were expressed only in testis, one (*ETY1*) showed intermediate expression pattern, two (*ETY2, ETY4*) were expressed ubiquitously, whereas *ETY3* did not show any expression in the panel of nine normal adult body tissues. ESTs *ETSTY1-6* were used for further studies as potential candidate genes for stallion fertility due to their multicopy nature and testis-specific expression. Four of these transcripts (*ETSTY3-6*)

demonstrated differential expression between fertile and infertile/subfertile individuals and are considered as strong candidate genes for stallion fertility.

The Y-linked novel transcripts are shown to be associated with spermatogenesis and other male fertility related function in human and mouse. For example, human *DAZ* (Deleted in Azoospermia) is a novel Y-linked gene cluster. Complete deletion of *DAZ* causes azoospermia (Yen et al. 1997), while three different patterns of partial *DAZ* deletions are associated with severe oligozoospermia and infertility (A et al. 2006). *DAZ* gene transcripts are shown to be localized in primary spermatocytes and *DAZ* gene activity seems to correspond to the proliferative activity of stem cells of germinal epithelium in patients with spermatogenic arrest (Szczerba et al. 2006). These findings strongly support the association of novel Y-linked genes (*DAZ* cluster) with spermatogenesis and thus, male fertility in human. Murine *Sly (Sycp3*-like Y-linked) is a novel multicopy gene on MMUY, is abundantly expressed in mouse spermatids and deletion of this gene may contribute to the abnormal sperm head development (Toure et al. 2005). *Sly* may also have a role in the development of acrosome and the regulation of gene expression in spermatids. *Sly* deletions lead to spermatogenic defects and male infertility in mouse (Reynard et al. 2009). Moreover, recent studies identified two novel Y genes in mouse, *AK006152* and *H2al2y*. Both are expressed specifically in spermatids indicating their role in mid to late spermatid development. Overexpression of these novel MMUY genes can potentially cause abnormal sperm head development and hence contribute to male infertility (Ferguson et al. 2009). To date, three multicopy novel genes (*FLJ36031Ya, TETY1* and *TETY2*) have been identified in cat Y chromosome

(Murphy et al. 2006). All three are expressed exclusively in cat testis and though the function of these genes is not yet clear, their testis-specific expression and multicopy nature indicates likely involvement in feline male fertility.

Findings in human, mouse and cat strongly suggest that novel, testis-specific genes in the horse Y chromosome (e. g., *ETSTY1-6*) also carry out important male fertility related functions. However, molecular details and genetic mechanisms of these functions remain topics for future studies. Partial sequence similarity between *ETSTY2*, *ETSTY3* and *ETSTY6* questions the uniqueness of these sequences – do they represent different regions of the same gene or are they transcripts of different genes. Based on sequence analysis using Sequencher V1.7 software we infer that the three sequences are not parts of a single transcript because at higher stringency parameters (90% and 20 nucleotides) there is no alignment between them. However, sequence similarity at lower stringency (<70% and 10 nucleotides) suggests that *ETSTY2*, *ETSTY3* and *ETSTY6* might be members of the same gene family. This is in agreement with the results of STS content analysis showing that *ETSTY2* and *ETSTY3* map to the same BAC clones in the multiocopy region in the distal part of Contig1 (Fig. A1). However, *ETSTY6* maps to different BAC clones and is located in the heterochromatic region which is several megabases proximal to the multicopy region. It is anticipated that generation of full-length cDNA sequences for these transcripts or complete sequencing of the horse MSY will resolve the relationships between the three sequences. Otherwise, the presence of gene families is a typical feature of the mammalian Y chromosome. For example, human *TSPY* gene family contains a cluster of genes: *TSPY1*, *TSPY2*, *TSPY3* and *TSPYL* (*TSPY-*

Like) which are not identical but share some sequence homology with each other and most likely carry out similar functions. The presence of novel, species-specific and male fertility related gene families, like *DAZ* in humans (Skaletsky et al. 2003) and *Sly* in mouse (Reynard et al. 2009), seems to be a characteristic feature of Y chromosomes. Our findings comply with these facts and show that such novel gene families are present also on the horse Y chromosome.

This study demonstrates the discovery and analysis of potential Y-linked candidate genes for male fertility in horses. First, transcriptional analysis of all Y-linked genes on a panel of body tissues identified those genes that are expressed specifically in testis. Next, qRT-PCR studies showed that some of these genes are differentially expressed in normal stallions compared to stallions with fertility problems. Gene copy number analysis showed that there is a correlation between gene expression level and copy numbers and that is associated with stallion fertility. Finally, full-length, nearly full length or partial cDNA sequences were obtained for several multicopy testis-specific transcripts and analyzed for their protein coding abilities. Taken together, besides human and mouse this is the first systematic search for Y-linked male fertility genes in mammals and the first of its kind in domestic species.

CHAPTER IV

CONCLUSIONS AND FUTURE WORK

CONCLUSIONS

This research represents the first systematic discovery, mapping and functional analysis of Y chromosome genes and ESTs in the horse. Although high-, medium- or low-resolution Y chromosome maps are available for a number of mammalian species, *viz*., human (Skaletsky et al. 2003), mouse (Mazeyrat et al. 1998), chimpanzee (Hughes et al. 2005; Kuroki et al. 2006), cattle (Liu et al. 2002), pig (Quilter et al. 2002), cat (Murphy et al. 1999a; Murphy et al. 2006; Pearks Wilkerson et al. 2008) and dog (Guyon et al. 2003a; Guyon et al. 2003b), direct association between Y-linked genes and male fertility has been shown so far only in humans and mice (Elliott et al. 1998; Matzuk and Lamb 2002; Skaletsky et al. 2003; Delbridge et al. 2004; Szczerba et al. 2006; Carrell 2008; Matzuk and Lamb 2008). Therefore, the present study is the first where the expression profiles, transcription levels and copy numbers of Y-linked genes have been related to male fertility in a species other than human and mouse. Given that stallion fertility is an important concern for the horse industry, the importance of the present findings cannot be underestimated. Although the research has focused on only a small fraction of the equine genome, it is an essential contribution to our understanding about the genetic component of male fertility in horses. Besides, detailed knowledge about the organization, gene content and functional profiles of the horse Y chromosome provides critical comparative information about the evolution of the sex chromosomes in

mammals. Finally, the high-resolution contig map of equine MSY is a foundation for complete sequencing of the horse Y chromosome – thus adding missing data to the female-based horse whole-genome-sequencing project.

FUTURE WORK

The comprehensive map of horse MSY presented in this study and the detailed map of the pseudoautosomal region (PAR), (Raudsepp and Chowdhary 2008) provide valuable resource for clone-based complete sequencing of the horse Y chromosome. Furthermore, the 197 BAC clones and 318 linearly ordered genes and STS markers in the contig map will serve as important landmarks for sequence assembly. Such landmarks are particularly essential for assembling Y chromosome sequences which are known to contain gene families, repetitive sequences, segmental duplications and palindromes (Skaletsky et al. 2003).

Complete sequencing of ECAY euchromatic region will aid the discovery of new Y-linked genes and will provide detailed information about the organization of the known genes. Knowledge about gene structure is necessary to identify regulatory elements, evaluate protein coding abilities and relate the structure to potential functions of these genes. Complete sequence data will also be instrumental for determining precise copy numbers of ECAY genes, identifying members of gene families and discovering copy number variations between individuals – both within normal population and between fertile and reproductively abnormal animals. Identification of minor sequence differences between individual copies of multicopy sequences will provide tools to

determine by PCR Y chromosome microdeletions or other rearrangements in the multicopy region of ECAY.

It is equally important to continue and expand the collection of samples (tissue, RNA, DNA) from stallions with a broad range of subfertile/infertile phenotypes. Only this way it will be possible to find out, for example, whether equine *TSPY* is indeed associated with sperm structural integrity or which Y-linked genes cause azoospermia and how. Larger number of samples is needed also to understand the role of other Y-linked genes - those which are not multicopy, testis-specific and novel. For example, it is not clear what kind of fertility related functions, if any, are carried out by X-degenerate genes or the genes that have been recruited to the Y chromosome from mitochondrial genome or autosomes.

The collection of ECAY genes, ESTs and STSs identified in this study will be a small but unique addition to equine whole genome analysis tools. Genes and ESTs will add Y-chromosome component to the gene expression arrays. The currently available platforms - the Texas A&M 21,000 element oligoarray (Chowdhary unpublished) and the Affymetrix 12,000 element oligonucleotide GeneChip (Nixon et al. 2008) - do not contain any expressed sequences from the Y chromosome. Sequences of Y-linked genes and ESTs will also be essential for the construction of custom-made specialized oligoarrays to study differential expression of sex and reproduction related (SRR) genes in horses and identify key-role male fertility genes genome-wide. Finally, with the horse genome sequenced and draft assembly available, it is anticipated that like in other species, whole genome tiling arrays will soon be constructed also for the horse. Y

chromosome sequences will, thus, be an important addition to the otherwise female-based genomic sequence data.

ECAY multicopy genes and variations in their copy numbers between normal individuals and between different fertility phenotypes is of particular interest. Recent studies in human and mouse demonstrate that genes present in regions of copy number variations (CVNs) are expressed in lower and in a more variable fashion than genes mapped elsewhere in the genome and can have a global influence on the transcriptome (Henrichsen et al. 2009). It has been proposed that copy number changes might influence gene expression through the perturbation of transcript structure (Reymond et al. 2007). CNVs have also been associated with various diseases or disease susceptibility phenotypes (Breunis et al. 2008; Hollox et al. 2008). It will, thus, be interesting to find out to which extent these properties of global CVNs apply to the multicopy genes/sequences in the horse Y chromosome.

Taken together, the findings of this study add new and important information to the current understanding of the genetic component of stallion and mammalian male fertility. Identification and characterization of a number of potential male fertility genes takes us a step closer to the ultimate goal of the genetics of stallion fertility– identify key-role genes, causative mutations and develop molecular diagnostic tests for early detection of subfertile/infertile individuals, so that owners and breeders can make informed decisions.

REFERENCES

A ZC, Yang Y, Zhang SZ, Lin L: Study on DAZ gene copy deletion in severe oligozoospermia sperm donor for ICSI. Yi Chuan 28:1057-1060 (2006).

Affara NA, Mitchell MJ: The role of human and mouse Y chromosome genes in male infertility. Journal of Endocrinological Investigation 23:630-645 (2000).

Affara NA: The role of the Y chromosome in male infertility. Exp Rev Mol Med (2001).

Agoulnik AI, Feng S: The genetics of cryptorchidism, in Carrell DT (ed): The Genetics of Male Fertility, pp 185-190 (Humana Press, Totowa, New Jersey 2007).

Agulnik AI, Bishop CE, Lerner JL, Agulnik SI, Solovyev VV: Analysis of mutation rates in the SMCY/SMCX genes shows that mammalian evolution is male driven. Mamm Genome 8:134-138 (1997).

Akimoto C, Kitagawa H, Matsumoto T, Kato S: Spermatogenesis-specific association of SMCY and MSH5. Genes Cells 13:623-633 (2008).

Amann RP, Veeramachaneni DN: Cryptorchidism in common eutherian mammals. Reproduction 133:541-561 (2007).

Bar-Joseph Z, Siegfried Z, Brandeis M, Brors B, Lu Y, Eils R, Dynlacht BD, Simon I: Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. Proc Natl Acad Sci USA 105:955-960 (2008).

Barthold JS: Undescended testis: current theories of etiology. Curr Opin Urol 18:395-400 (2008).

Beardsley A, O'Donnell L: Characterization of normal spermiation and spermiation failure induced by hormone suppression in adult rats. Biol Reprod 68:1299-1307 (2003).

Betran E, Emerson JJ, Kaessmann H, Long M: Sex chromosomes and male functions: where do new genes go? Cell Cycle 3:873-875 (2004).

Birnboim HC, Doly J: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513-1523 (1979).

Birnboim HC: A rapid alkaline extraction method for the isolation of plasmid DNA. Methods Enzymol 100:243-255 (1983).

Blanchard TL, Varner DD: Evaluating breeding soundness in stallions. Vet Med 815-818 (1997).

Boettger-Tong HL, Agulnik AI, Ty TI, Bishop CE: Transposition of RhoA to the murine Y chromosome. Genomics 49:180-187 (1998).

Bosard T, Love C, Brinsko S, Blanchard T, Thompson J, Varner D: Evaluation and diagnosis of acrosome function/dysfunction in the stallion. Anim Reprod Sci 89:215-217 (2005).

Braun RE: Packaging paternal chromosomes with protamine. Nat Genet 28:10-12 (2001).

Breunis WB, van Mirre E, Bruin M, Geissler J, de Boer M, Peters M, Roos D, de Haas M, Koene HR, Kuijpers TW: Copy number variation of the activating FCGR2C gene predisposes to idiopathic thrombocytopenic purpura. Blood 111:1029-1038 (2008).

Bullejos M, Sanchez A, Burgos M, Jimenez R, Diaz De La Guardia R: Multiple mono- and polymorphic Y-linked copies of the SRY HMG-box in microtidae. Cytogenet Cell Genet 86:46-50 (1999).

Burgoyne PS: Genetic homology and crossing over in the X and Y chromosomes of Mammals. Hum Genet 61:85-90 (1982).

Canela N, Rodriguez-Vilarrupla A, Estanyol JM, Diaz C, Pujol MJ, Agell N, Bachs O: The SET protein regulates G2/M transition by modulating cyclin B-cyclin-dependent kinase 1 activity. J Biol Chem 278:1158-1164 (2003).

Carrell DT: The genetics of male infertility in the era of genomics, in Carrell DT (ed): The Genetics of Male Infertility, pp 3-27 (Humana Press, Totowa, New Jersey 2007).

Carrell DT: The clinical implementation of sperm chromosome aneuploidy testing: pitfalls and promises. J Androl 29:124-133 (2008).

Chai Z, Sarcevic B, Mawson A, Toh BH: SET-related cell division autoantigen-1 (CDA1) arrests cell growth. J Biol Chem 276:33665-33674 (2001).

Chandley AC, Cooke HJ: Human male fertility--Y-linked genes and spermatogenesis. Hum Mol Genet 3 Spec No:1449-1452 (1994).

Charlesworth B: The evolution of sex chromosomes. Science 251:1030-1033 (1991).

Charlesworth B, Charlesworth D: The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci 355:1563-1572 (2000).

Charlesworth B: The evolution of chromosomal sex determination. Novartis Found Symp 244:207-219; discussion 220-204, 253-207 (2002).

Charlesworth D, Charlesworth B: Sex chromosomes: evolution of the weird and wonderful. Curr Biol 15:R129-131 (2005).

Chen-Liu LW, Huang BC, Scalzi JM, Hall BK, Sims KR, Davis LM, Siebert PD, Hozier JC: Selection of hybrids by affinity capture (SHAC): a method for the generation of cDNAs enriched in sequences from a specific chromosome region. Genomics 30:388-392 (1995).

Cheng CY, Mruk DD: Cell junction dynamics in the testis: Sertoli-germ cell interactions and male contraceptive development. Physiol Rev 82:825-874 (2002).

Cheng H, Shi H, Zhou R, Guo Y, Liu L, Liu J, Jiang Y, Kudo T, Sutou S: Characterization of bovidae sex-determining gene SRY. Genet Sel Evol 33:687-694 (2001).

Chenoweth PJ: Genetic sperm defects. Theriogenology 64:457-468 (2005).

Chowdhary BP, Raudsepp T, Kata SR, Goh G, Millon LV, Allan V, Piumi F, Guerin G, Swinburne J, Binns M, Lear TL, Mickelson J, Murray J, Antczak DF, Womack JE, Skow LC: The first-generation whole-genome radiation hybrid map in the horse identifies conserved segments in human and mouse genomes. Genome Res 13:742-751 (2003).

Colenbrander B, Gadella BM, Stout TA: The predictive value of semen analysis in the evaluation of stallion fertility. Reprod Domest Anim 38:305-311 (2003).

Compagnone NA, Zhang P, Vigne JL, Mellon SH: Novel role for the nuclear phosphoprotein SET in transcriptional activation of P450c17 and initiation of neurosteroidogenesis. . Mol Endocrinol 14:875–888 (2000).

Conner SJ, Lefievre L, Kirkman-Brown J, Machado-Oliveira GSM, Michelangeli F, Publicover SJ, Barratt CLR: Physiological and proteomic approaches to understanding human sperm function, in Carrell DT (ed): The Genetics of Male Infertility, pp 77-97 (Humana Press, Totowa, New Jersey 2007).

Cooke HJ, Saunders PT: Mouse models of male infertility. Nat Rev Genet 3:790-801 (2002).

Crew FA: A suggestion as to the cause of the aspermatic condition of the imperfectly descended testis. J Anat 56:98-106 (1922).

Dada R, Gupta NP, Kucheria K: Yq microdeletions--azoospermia factor candidate genes and spermatogenic arrest. J Biomol Tech 15:176-183 (2004).

Daggag H, Svingen T, Western PS, van den Bergen JA, McClive PJ, Harley VR, Koopman P, Sinclair AH: The Rhox homeobox gene family shows sexually dimorphic and dynamic expression during mouse embryonic gonad development. Biol Reprod 79:468-474 (2008).

Dam AH, Feenstra I, Westphal JR, Ramos L, van Golde RJ, Kremer JA: Globozoospermia revisited. Hum Reprod Update 13:63-75 (2007).

De Blas GA, Roggero CM, Tomes CN, Mayorga LS: Dynamics of SNARE assembly and disassembly during sperm acrosomal exocytosis. PLoS Biol 3:e323 (2005).

Dechend F, Schubert S, Nanda I, Vogel T, Schmid M, Schmidtke J: Organization and expression of rat Tspy. Cytogenet Cell Genet 83:270-274 (1998).

Del Mastro RG, Lovett M: Isolation of coding sequences from genomic regions using direct selection. Methods Mol Biol 68:183-199 (1997).

Delbridge ML, Longepied G, Depetris D, Mattei MG, Disteche CM, Marshall Graves JA, Mitchell MJ: TSPY, the candidate gonadoblastoma gene on the human Y chromosome, has a widely expressed homologue on the X - implications for Y chromosome evolution. Chromosome Res 12:345-356 (2004).

Delbridge ML, Graves JA: Origin and evolution of spermatogenesis genes on the human sex chromosomes. Soc Reprod Fertil Suppl 65:1-17 (2007).

Dieterich K, Zouari R, Harbuz R, Vialard F, Martinez D, Bellayou H, Prisant N, Zoghmar A, Guichaoua MR, Koscinski I, Kharouf M, Noruzinia M, Nadifi S, Sefiani A, Lornage J, Zahi M, Viville S, Sele B, Jouk PS, Jacob MC, Escalier D, Nikas Y, Hennebicq S, Lunardi J, Ray PF: The Aurora Kinase C c.144delC mutation causes meiosis I arrest in men and is frequent in the North African population. Hum Mol Genet 18:1301-1309 (2009).

Ditton HJ, Zimmer J, Kamp C, Rajpert-De Meyts E, Vogt PH: The AZFa gene DBY (DDX3Y) is widely transcribed but the protein is limited to the male germ cells by translation control. Hum Mol Genet 13:2333-2341 (2004).

Dorus S, Gilbert SL, Forster ML, Barndt RJ, Lahn BT: The CDY-related gene family: coordinated evolution in copy number, expression profile and protein sequence. Hum Mol Genet 12:1643-1650 (2003).

Elliott DJ, Ma K, Kerr SM, Thakrar R, Speed R, Chandley AC, Cooke H: An RBM homologue maps to the mouse Y chromosome and is expressed in germ cell. Hum Mol Genet 5:869–874 (1996).

Elliott DJ, Millar MR, Oghene K, Ross A, Kiesewetter F, Pryor J, McIntyre M, Hargreave TB, Saunders PTK, Vogt PH, Chandley AC, Cooke H: Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. . Proc Natl Acad Sci USA 94:3848–3853 (1997).

Elliott DJ, Oghene K, Makarov G, Makarova O, Hargreave TB, Chandley AC, Eperon IC, Cooke HJ: Dynamic changes in the subnuclear organisation of pre-mRNA splicing proteins and RBM during human germ cell development. . J Cell Sci 111:1255–1265 (1998).

Ellis PJ, Clemente EJ, Ball P, Toure A, Ferguson L, Turner JM, Loveland KL, Affara NA, Burgoyne PS: Deletions on mouse Yq lead to upregulation of multiple X- and Y-linked transcripts in spermatids. Hum Mol Genet 14:2705-2715 (2005).

Ellis PJ, Affara NA: Spermatogenesis and sex chromosome gene content: an evolutionary perspective. Hum Fertil (Camb) 9:1-7 (2006).

Ellis PJ, Ferguson L, Clemente EJ, Affara NA: Bidirectional transcription of a novel chimeric gene mapping to mouse chromosome Yq. BMC Evol Biol 7:171 (2007).

Emerson JJ, Kaessmann H, Betran E, Long M: Extensive gene traffic on the mammalian X chromosome. Science 303:537-540 (2004).

Feng CQ, Song YB, Zou YG, Mao XM: Mutation of MTCYB and MTATP6 is associated with asthenospermia. Zhonghua Nan Ke Xue 14:321-323 (2008).

Feng S, Cortessis VK, Hwang A, Hardy B, Koh CJ, Bogatcheva NV, Agoulnik AI: Mutation analysis of INSL3 and GREAT/LGR8 genes in familial cryptorchidism. Urology 64:1032-1036 (2004).

Ferguson L, Ellis PJ, Affara NA: Two novel mouse genes mapped to chromosome Yp are expressed specifically in spermatids. Mamm Genome 20:193-206 (2009).

Ferlin A, Moro E, Rossi A, Dallapiccola B, Foresta C: The human Y chromosome's azoospermia factor b (AZFb) region: sequence, structure, and deletion analysis in infertile men. J Med Genet 40:18-24 (2003).

Filipponi D, Feil R: Perturbation of genomic imprinting in oligozoospermia. Epigenetics 4:27-30 (2009).

Foresta C, Zuccarello D, Garolla A, Ferlin A: Role of hormones, genes, and environment in human cryptorchidism. Endocr Rev 29:560-580 (2008).

Freije D, Helms C, Watson MS, Donis-Keller H: Identification of a second pseudoautosomal region near the Xq and Yq telomeres. Science 258:1784-1787 (1992).

Fukami K, Nakao K, Inoue T, Kataoka Y, Kurokawa M, Fissore RA, Nakamura K, Katsuki M, Mikoshiba K, Yoshida N, Takenawa T: Requirement of phospholipase Cdelta4 for the zona pellucida-induced acrosome reaction. Science 292:920-923 (2001).

Geraldes A, Rogel-Gaillard C, Ferrand N: High levels of nucleotide diversity in the European rabbit (*Oryctolagus cuniculus*) SRY gene. Anim Genet 36:349-351 (2005).

Geraldes A, Ferrand N: A 7-bp insertion in the 3' untranslated region suggests the duplication and concerted evolution of the rabbit SRY gene. Genet Sel Evol 38:313-320 (2006).

Giachini C, Guarducci E, Longepied G, Degl'Innocenti S, Becherini L, Forti G, Mitchell MJ, Krausz C: The gr/gr deletion(s): a new genetic test in male infertility? J Med Genet 42:497-502 (2005).

Gibson UE, Heid CA, Williams PM: A novel method for real time quantitative RT-PCR. Genome Res 6:995-1001 (1996).

Gonsalves J, Sun F, Schlegel PN, Turek PJ, Hopps CV, Greene C, Martin RH, Pera RA: Defective recombination in infertile men. Hum Mol Genet 13:2875-2883 (2004).

Goodfellow PJ, Darling SM, Thomas NS, Goodfellow PN: A pseudoautosomal gene in man. Science 234:740-743 (1986).

Goto H, Peng L, Makova KD: Evolution of X-degenerate Y chromosome genes in greater apes: conservation of gene content in human and gorilla, but not chimpanzee. J Mol Evol 68:134-144 (2009).

Graham JK: Analysis of stallion semen and its relation to fertility. Vet Clin North Am Equine Pract 12:119-130 (1996).

Graves JA: The evolution of mammalian sex chromosomes and the origin of sex determining genes. Philos Trans R Soc Lond B Biol Sci 350:305-311; discussion 311-302 (1995).

Graves JA: Evolution of the mammalian Y chromosome and sex-determining genes. J Exp Zool 281:472-481 (1998).

Graves JA: The degenerate Y chromosome - can conversion save it? Reprod Fertil Dev 16:527-534 (2004).

Graves JA: Sex chromosome specialization and degeneration in mammals. Cell 124:901-914 (2006).

Graves JA, Koina E, Sankovic N: How the gene content of human sex chromosomes evolved. Curr Opin Genet Dev 16:219-224 (2006).

Grzmil P, Golas A, Muller C, Styrna J: The influence of the deletion on the long arm of the Y chromosome on sperm motility in mice. Theriogenology 67:760-766 (2007).

Guan XY, Xu J, Anzick SL, Zhang H, Trent JM, Meltzer PS: Hybrid selection of transcribed sequences from microdissected DNA: isolation of genes within amplified region at 20q11-q13.2 in breast cancer. Cancer Res 56:3446-3450 (1996).

Guescini M, Sisti D, Rocchi MB, Stocchi L, Stocchi V: A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition. BMC Bioinformatics 9:326 (2008).

Gustafson AL, Tallmadge RL, Ramlachan N, Miller D, Bird H, Antczak DF, Raudsepp T, Chowdhary BP, Skow LC: An ordered BAC contig map of the equine major histocompatibility complex. Cytogenet Genome Res 102:189-195 (2003).

Guyon R, Kirkness EF, Lorentzen TD, Hitte C, Comstock KE, Quignon P, Derrien T, Andre C, Fraser CM, Galibert F, Ostrander EA: Building comparative maps using 1.5x sequence coverage: human chromosome 1p and the canine genome. Cold Spring Harb Symp Quant Biol 68:171-177 (2003a).

Guyon R, Lorentzen TD, Hitte C, Kim L, Cadieu E, Parker HG, Quignon P, Lowe JK, Renier C, Gelfenbeyn B, Vignaux F, DeFrance HB, Gloux S, Mahairas GG, Andre C, Galibert F, Ostrander EA: A 1-Mb resolution radiation hybrid map of the canine genome. Proc Natl Acad Sci USA 100:5296-5301 (2003b).

Gvozdev VA, Kogan GL, Usakin LA: The Y chromosome as a target for acquired and amplified genetic material in evolution. Bioessays 27:1256-1262 (2005).

Hall NM, Brown GM, Furlong RA, Sargent CA, Mitchell M, Rocha D, Affara NA: Usp9y (ubiquitin-specific protease 9 gene on the Y) is associated with a functional promoter and encodes an intact open reading frame homologous to Usp9x that is under selective constraint. Mamm Genome 14:437-447 (2003).

Hamann H, Jude R, Sieme H, Mertens U, Topfer-Petersen E, Distl O, Leeb T: A polymorphism within the equine CRISP3 gene is associated with stallion fertility in Hanoverian warmblood horses. Anim Genet 38:259-264 (2007).

Hargreave TB: Genetics and male infertility. Curr Opin Obstet Gynecol 12:207-219 (2000).

Hasegawa T, Ishida M, Harigaya T, Sato F, Ishida N, Mukoyama H: Linear SRY transcript in equine testis. J Vet Med Sci 61:97-100 (1999).

Heid CA, Stevens J, Livak KJ, Williams PM: Real time quantitative PCR. Genome Res 6:986-994 (1996).

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J: qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 8:R19 (2007).

Henrichsen CN, Chaignat E, Reymond A: Copy number variants, diseases and gene expression. Hum Mol Genet 18:R1-8 (2009).

Hiramatsu R, Matoba S, Kanai-Azuma M, Tsunekawa N, Katoh-Fukui Y, Kurohmaru M, Morohashi K, Wilhelm D, Koopman P, Kanai Y: A critical time window of Sry action in gonadal sex determination in mice. Development 136:129-138 (2009).

Hirota K, Piumi F, Sato F, Ishida N, Guerin G, Miura N, Hasegawa T: FISH assignment of two equine BAC clones containing SRY and ZFY. Anim Genet 32:326-327 (2001).

Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, Rodijk-Olthuis D, van de Kerkhof PC, Traupe H, de Jongh G, den Heijer M, Reis A, Armour JA, Schalkwijk J: Psoriasis is associated with increased beta-defensin genomic copy number. Nat Genet 40:23-25 (2008).

Honecker F, Stoop H, de Krijger RR, Chris Lau YF, Bokemeyer C, Looijenga LH: Pathobiological implications of the expression of markers of testicular carcinoma in situ by fetal germ cells. J Pathol 203:849-857 (2004).

Hughes JF, Skaletsky H, Pyntikova T, Minx PJ, Graves T, Rozen S, Wilson RK, Page DC: Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. Nature 437:100-103 (2005).

Hughes V: Geneticists crack the code of infertility. Nat Med 14:1174 (2008).

Iguchi N, Yang S, Lamb DJ, Hecht NB: An SNP in protamine 1: a possible genetic cause of male infertility? J Med Genet 43:382-384 (2006).

Ing NH, Laughlin AM, Varner DD, Welsh TH, Jr., Forrest DW, Blanchard TL, Johnson L: Gene expression in the spermatogenically inactive "dark" and maturing "light" testicular tissues of the prepubertal colt. J Androl 25:535-544 (2004).

Jin JL, O'Doherty AM, Wang S, Zheng H, Sanders KM, Yan W: Catsper3 and catsper4 encode two cation channel-like proteins exclusively expressed in the testis. Biol Reprod 73:1235-1242 (2005).

Kim Y, Kobayashi A, Sekido R, DiNapoli L, Brennan J, Chaboissier MC, Poulat F, Behringer RR, Lovell-Badge R, Capel B: Fgf9 and Wnt4 act as antagonistic signals to regulate mammalian sex determination. PLoS Biol 4:e187 (2006).

Kimmins S, Sassone-Corsi P: Chromatin remodelling and epigenetic features of germ cells. Nature 434:583-589 (2005).

Kirsch S, Weiss B, Miner TL, Waterston RH, Clark RA, Eichler EE, Munch C, Schempp W, Rappold G: Interchromosomal segmental duplications of the pericentromeric region on the human Y chromosome. Genome Res 15:195-204 (2005).

Kirsch S, Munch C, Jiang Z, Cheng Z, Chen L, Batz C, Eichler EE, Schempp W: Evolutionary dynamics of segmental duplications from human Y-chromosomal euchromatin/heterochromatin transition regions. Genome Res 18:1030-1042 (2008).

Kitamura K, Tanaka H, Nishimune Y: The RING-finger protein haprin: domains and function in the acrosome reaction. Curr Protein Pept Sci 6:567-574 (2005).

Klonisch T, Fowler PA, Hombach-Klonisch S: Molecular and genetic regulation of testis descent and external genitalia development. Dev Biol 270:1-18 (2004).

Kostiner DR, Turek PJ, Reijo RA: Male infertility: analysis of the markers and genes on the human Y chromosome. Hum Reprod 13:3032-3038 (1998).

Krausz C, Degl'Innocenti S, Nuti F, Morelli A, Felici F, Sansone M, Varriale G, Forti G: Natural transmission of USP9Y gene mutations: a new perspective on the role of AZFa genes in male fertility. Hum Mol Genet 15:2673-2681 (2006).

Krausz C, Giachini C: Genetic risk factors in male infertility. Arch Androl 53:125-133 (2007).

Kuroki Y, Toyoda A, Noguchi H, Taylor TD, Itoh T, Kim DS, Kim DW, Choi SH, Kim IC, Choi HH, Kim YS, Satta Y, Saitou N, Yamada T, Morishita S, Hattori M, Sakaki Y, Park HS, Fujiyama A: Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway. Nat Genet 38:158-167 (2006).

Kvaloy K, Galvagni F, Brown WR: The sequence organization of the long arm pseudoautosomal region of the human sex chromosomes. Hum Mol Genet 3:771-778 (1994).

Lahn BT, Page DC: Functional coherence of the human Y chromosome. Science 278:675-680 (1997).

Lahn BT, Page DC: Four evolutionary strata on the human X chromosome. Science 286:964-967 (1999).

Lahn BT, Page DC: A human sex-chromosomal gene family expressed in male germ cells and encoding variably charged proteins. Hum Mol Genet 9:311-319 (2000).

Lahn BT, Pearson NM, Jegalian K: The human Y chromosome, in the light of evolution. Nat Rev Genet 2:207-216 (2001).

Lange J, Skaletsky H, Bell GW, Page DC: MSY Breakpoint Mapper, a database of sequence-tagged sites useful in defining naturally occurring deletions in the human Y chromosome. Nucleic Acids Res 36:D809-814 (2008).

Lardone MC, Parodi DA, Ebensperger M, Penaloza P, Cornejo V, Valdevenito R, Pommer R, Castro A: AZFc partial deletions in Chilean men with severe spermatogenic failure. Fertil Steril 88:1318-1326 (2007a).

Lardone MC, Parodi DA, Valdevenito R, Ebensperger M, Piottante A, Madariaga M, Smith R, Pommer R, Zambrano N, Castro A: Quantification of DDX3Y, RBMY1, DAZ and TSPY mRNAs in testes of patients with severe impairment of spermatogenesis. Mol Hum Reprod 13:705-712 (2007b).

Lau YF: Gonadoblastoma, testicular and prostate cancers, and the TSPY gene. Am J Hum Genet 64:921-927 (1999).

Lau YF, Lau HW, Komuves LG: Expression pattern of a gonadoblastoma candidate gene suggests a role of the Y chromosome in prostate cancer. Cytogenet Genome Res 101:250-260 (2003).

Lau YF, Li Y, Kido T: Gonadoblastoma locus and the TSPY gene on the human Y chromosome. Birth Defects Res C Embryo Today 87:114-122 (2009).

Lee KF, Tam YT, Zuo Y, Cheong AW, Pang RT, Lee NP, Shum CK, Tam PC, Cheung AN, Yang ZM, Yeung WS, Luk JM: Characterization of an acrosome protein VAD1.2/AEP2 which is differentially expressed in spermatogenesis. Mol Hum Reprod 14:465-474 (2008).

Lee KH, Song GJ, Kang IS, Kim SW, Paick JS, Chung CH, Rhee K: Ubiquitin-specific protease activity of USP9Y, a male infertility gene on the Y chromosome. Reprod Fertil Dev 15:129-133 (2003).

Leeb T, Sieme H, Topfer-Petersen E: Genetic markers for stallion fertility--lessons from humans and mice. Anim Reprod Sci 89:21-29 (2005).

Lessard C, Lothrop H, Schimenti JC, Handel MA: Mutagenesis-generated mouse models of human infertility with abnormal sperm. Hum Reprod 22:159-166 (2007).

Levy N, Navarro A, Bishop CE, Mitchell MJ: The ubiquitin-activating enzyme E1 homologous genes on the mouse Y chromosome (Ube1y) represent one functional gene and six partial pseudogenes. Mamm Genome 11:164-168 (2000).

Li L, Hamer DH: Recombination and allelic association in the Xq/Yq homology region. Hum Mol Genet 4:2013-2016 (1995).

Lindgren G, Backstrom N, Swinburne J, Hellborg L, Einarsson A, Sandberg K, Cothran G, Vila C, Binns M, Ellegren H: Limited number of patrilines in horse domestication. Nat Genet 36:335-336 (2004).

Liu WS, Mariani P, Beattie CW, Alexander LJ, Ponce De Leon FA: A radiation hybrid map for the bovine Y Chromosome. Mamm Genome 13:320-326 (2002).

Liu WS, Beattie CW, Ponce de Leon FA: Bovine Y chromosome microsatellite polymorphisms. Cytogenet Genome Res 102:53-58 (2003).

Liu WS, de Leon FA: Assignment of SRY, ANT3, and CSF2RA to the bovine Y chromosome by FISH and RH mapping. Anim Biotechnol 15:103-109 (2004).

Lochamy J, Rogers EM, Boss JM: CREB and phospho-CREB interact with RFX5 and CIITA to regulate MHC class II genes. Mol Immunol 44:837-847 (2007).

Lopez CI, Belmonte SA, De Blas GA, Mayorga LS: Membrane-permeant Rab3A triggers acrosomal exocytosis in living human sperm. FASEB J 21:4121-4130 (2007).

Love CC, Kenney RM: The relationship of increased susceptibility of sperm DNA to denaturation and fertility in the stallion. Theriogenology 50:955-972 (1998).

Love CC, Thompson JA, Brinsko SP, Rigby SL, Blanchard TL, Lowry VK, Varner DD: Relationship between stallion sperm motility and viability as detected by two fluorescence staining techniques using flow cytometry. Theriogenology 60:1127-1138 (2003).

Lovett M, Kere J, Hinton LM: Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. Proc Natl Acad Sci U S A 88:9628-9632 (1991).

Luddi A, Margollicci M, Gambera L, Serafini F, Cioni M, De Leo V, Balestri P, Piomboni P: Spermatogenesis in a man with complete deletion of USP9Y. N Engl J Med 360:881-885 (2009).

Luk JM, Lee NP, Shum CK, Lam BY, Siu AF, Che CM, Tam PC, Cheung AN, Yang ZM, Lin YN, Matzuk MM, Lee KF, Yeung WS: Acrosome-specific gene AEP1: identification, characterization and roles in spermatogenesis. J Cell Physiol 209:755-766 (2006).

Luu-The V, Paquet N, Calvo E, Cumps J: Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction. Biotechniques 38:287-293 (2005).

MacLean JA, 2nd, Rao MK, Doyle KM, Richards JS, Wilkinson MF: Regulation of the Rhox5 homeobox gene in primary granulosa cells: preovulatory expression and dependence on SP1/SP3 and GABP. Biol Reprod 73:1126-1134 (2005).

Madill S: Reproductive considerations: mare and stallion. Vet Clin North Am Equine Pract 18:591-619 (2002).

Makrinou E, Fox M, Lovett M, Haworth K, Cameron JM, Taylor K, Edwards YH: TTY2: a multicopy Y-linked gene family. Genome Res 11:935-945 (2001).

Mangs AH, Morris BJ: The human pseudoautosomal region (PAR): origin, function and future. Curr Genomics 8:129–136 (2007).

Marengo SR: Maturing the sperm: unique mechanisms for modifying integral proteins in the sperm plasma membrane. Anim Reprod Sci 105:52-63 (2008).

Margarit E, Guillen A, Rebordosa C, Vidal-Taboada J, Sanchez M, Ballesta F, Oliva R: Identification of conserved potentially regulatory sequences of the SRY gene from 10 different species of mammals. Biochem Biophys Res Commun 245:370-377 (1998).

Matzuk MM, Lamb DJ: Genetic dissection of mammalian fertility pathways. Nat Cell Biol 4 Suppl:s41-49 (2002).

Matzuk MM, Lamb DJ: The biology of infertility: research advances and clinical challenges. Nat Med 14:1197-1213 (2008).

Mayorga LS, Tomes CN, Belmonte SA: Acrosomal exocytosis, a special type of regulated secretion. IUBMB Life 59:286-292 (2007).

Mazeyrat S, Mitchell MJ: Rodent Y chromosome TSPY gene is functional in rat and non-functional in mouse. Hum Mol Genet 7:557-562 (1998).

Mazeyrat S, Saut N, Sargent CA, Grimmond S, Longepied G, Ehrmann IE, Ellis PS, Greenfield A, Affara NA, Mitchell MJ: The mouse Y chromosome interval necessary for spermatogonial proliferation is gene dense with syntenic homology to the human AZFa region. Hum Mol Genet 7:1713-1724 (1998).

Mazeyrat S, Saut N, Mattei MG, Mitchell MJ: RBMY evolved on the Y chromosome from a ubiquitously transcribed X-Y identical gene. Nat Genet 22:224-226 (1999).

McElreavey K, Krausz C, Patrat C, Fellous M: Male infertility and microdeletions of the Y chromosome. Gynecol Obstet Fertil 30:405-412 (2002).

Meizel S, Son JH: Studies of sperm from mutant mice suggesting that two neurotransmitter receptors are important to the zona pellucida-initiated acrosome reaction. Mol Reprod Dev 72:250-258 (2005).

Merkies K, Buhr MM: Stallion spermatozoa viability: comparison of flow cytometry with other methods. Methods Mol Biol 253:49-58 (2004).

Milenkovic D, Martin P, Guerin G, Leroux C: A specific pattern of splicing for the horse alphaS1-Casein mRNA and partial genomic characterization of the relevant locus. Genet Sel Evol 34:509-519 (2002).

Mitchell MJ, Woods DR, Tucker PK, Opp JS, Bishop CE: Homology of a candidate spermatogenic gene from the mouse Y chromosome to the ubiquitin-activating enzyme E1. Nature 354:483-486 (1991).

Mitchell MJ, Wilcox SA, Watson JM, Lerner JL, Woods DR, Scheffler J, Hearn JP, Bishop CE, Graves JAM: The origin and loss of the ubiquitin activating enzyme gene on the mammalian Y chromosome. Human Molecular Genetics 7:429-434 (1998).

Mitchell MJ: Spermatogenesis and the mouse Y chromosome: specialisation out of decay. Results Probl Cell Differ 28:233-270 (2000).

Moore SS, Byrne K, Johnson SE, Kata S, Womack JE: Physical mapping of CSF2RA, ANT3 and STS on the pseudoautosomal region of bovine chromosome X. Anim Genet 32:102-104 (2001).

Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC, Turner JM: The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. Nat Genet 40:794-799 (2008).

Murphy WJ, Menotti-Raymond M, Lyons LA, Thompson MA, O'Brien SJ: Development of a feline whole genome radiation hybrid panel and comparative mapping of human chromosome 12 and 22 loci. Genomics 57:1-8 (1999a).

Murphy WJ, Sun S, Chen ZQ, Pecon-Slattery J, O'Brien SJ: Extensive conservation of sex chromosome organization between cat and human revealed by parallel radiation hybrid mapping. Genome Res 9:1223-1230 (1999b).

Murphy WJ, Pearks Wilkerson AJ, Raudsepp T, Agarwala R, Schaffer AA, Stanyon R, Chowdhary BP: Novel gene acquisition on carnivore Y chromosomes. PLoS Genet 2:e43 (2006).

Nagamine CM: The testis-determining gene, SRY, exists in multiple copies in Old World rodents. Genet Res 64:151-159 (1994).

Nagata K, Kawase H, Handa H, Yano K, Yamasaki M, Ishimi Y, Okuda A, Kikuchi A, Matsumoto K: Replication factor encoded by a putative oncogene, set, associated with myeloid leukemogenesis. . Proc Natl Acad Sci USA 92:4279– 4283 (1995).

Nayernia K, Meinhardt A, Drabent B, Adham IM, Muller C, Steckel M, Sancken U, Engel W: Synergistic effects of germ cell expressed genes on male fertility in mice. Cytogenet Genome Res 103:314-320 (2003).

Nef S, Verma-Kurvari S, Merenmies J, Vassalli JD, Efstratiadis A, Accili D, Parada LF: Testis determination requires insulin receptor family function in mice. Nature 426:291-295 (2003).

Neild DN, Gadella BM, Aguero A, Stout TA, Colenbrander B: Capacitation, acrosome function and chromatin structure in stallion sperm. Anim Reprod Sci 89:47-56 (2005).

Nixon AJ, Kerstin EG, Sun Q, Well ME: Development and characterization of a novel large-scale whole transcript olligonucleotide exon array for the horse. Poster Abstract W152: Plant and Animal Genome (PAG) San Diego, California 2008).

O'Bryan MK, de Kretser D: Mouse models for genes involved in impaired spermatogenesis. Int J Androl 29:76-89; discussion 105-108 (2006).

Obermann H, Raabe I, Balvers M, Brunswig B, Schulze W, Kirchhoff C: Novel testis-expressed profilin IV associated with acrosome biogenesis and spermatid elongation. Mol Hum Reprod 11:53-64 (2005).

Odorisio T, Mahadevaiah SK, McCarrey JR, Burgoyne PS: Transcriptional analysis of the candidate spermatogenesis gene Ube1y and of the closely related Ube1x shows that they are coexpressed in spermatogonia and spermatids but are repressed in pachytene spermatocytes. Dev Biol 180:336-343 (1996).

Ohno S: Sex Chromosomes and Sex-linked Genes., ed 1. (Springer-Verlag, New York 1967).

Olivier M, Breen M, Binns MM, Lust G: Localization and characterization of nucleotide sequences from the canine Y chromosome. Chromosome Res 7:223-233 (1999).

Pannetier M, Tilly G, Kocer A, Hudrisier M, Renault L, Chesnais N, Costa J, Le Provost F, Vaiman D, Vilotte JL, Pailhoux E: Goat SRY induces testis development in XX transgenic mice. FEBS Lett 580:3715-3720 (2006).

Parma P, Pailhoux E, Cotinot C: Reverse transcription-polymerase chain reaction analysis of genes involved in gonadal differentiation in pigs. Biol Reprod 61:741-748 (1999).

Payen E, Pailhoux E, Abou Merhi R, Gianquinto L, Kirszenbaum M, Locatelli A, Cotinot C: Characterization of ovine SRY transcript and developmental

expression of genes involved in sexual differentiation. Int J Dev Biol 40:567-575 (1996).

Pearks Wilkerson AJ, Raudsepp T, Graves T, Albracht D, Warren W, Chowdhary BP, Skow LC, Murphy WJ: Gene discovery and comparative analysis of X-degenerate genes from the domestic cat Y chromosome. Genomics 92:329-338 (2008).

Penedo MC, Millon LV, Bernoco D, Bailey E, Binns M, Cholewinski G, Ellis N, Flynn J, Gralak B, Guthrie A, Hasegawa T, Lindgren G, Lyons LA, Roed KH, Swinburne JE, Tozaki T: International equine gene mapping workshop report: A comprehensive linkage map constructed with data from new markers and by merging four mapping resources. Cytogenet Genome Res 111:5-15 (2005).

Perry GH, Tito RY, Verrelli BC: The evolutionary history of human and chimpanzee Y-chromosome gene loss. Mol Biol Evol 24:853–859 (2007).

Perry J, Palmer S, Gabriel A, Ashworth A: A short pseudoautosomal region in laboratory mice. Genome Res 11:1826-1832 (2001).

Peterlin B, Kunej T, Hristovski D: Diagnostic test for Y chromosome microdeletion screening in male infertility. Genet Test 8:45-49 (2004).

Pirrello O, Machev N, Schimdt F, Terriou P, Menezo Y, Viville S: Search for mutations involved in human globozoospermia. Hum Reprod 20:1314-1318 (2005).

Polanco JC, Koopman P: Sry and the hesitant beginnings of male development. Dev Biol 302:13-24 (2007).

Power MM: Y chromosome length variation and its significance in the horse. J Hered 79:311-313 (1988).

Prosser J, Inglis JD, Condie A, Ma K, Kerr S, Thakrar R, Taylor K, Cameron JM, Cooke HJ: Degeneracy in human multicopy RBM (YRRM), a candidate spermatogenesis gene. Mamm Genome 7:835-842 (1996).

Quilter CR, Blott SC, Mileham AJ, Affara NA, Sargent CA, Griffin DK: A mapping and evolutionary study of porcine sex chromosome genes. Mamm Genome 13:588-594 (2002).

Quintana-Murci L, Fellous M: The human Y chromosome: The biological role of a "Functional Wasteland". J Biomed Biotechnol 1:18-24 (2001).

Rappold GA: The pseudoautosomal regions of the human sex chromosomes. Hum Genet 92:315-324 (1993).

Raudsepp T, Chowdhary BP: Construction of chromosome-specific paints for meta- and submetacentric autosomes and the sex chromosomes in the horse and their use to detect homologous chromosomal segments in the donkey. Chromosome Res 7:103-114 (1999).

Raudsepp T, Lee EJ, Kata SR, Brinkmeyer C, Mickelson JR, Skow LC, Womack JE, Chowdhary BP: Exceptional conservation of horse-human gene order on X chromosome revealed by high-resolution radiation hybrid mapping. Proc Natl Acad Sci U S A 101:2386-2391 (2004a).

Raudsepp T, Santani A, Wallner B, Kata SR, Ren C, Zhang HB, Womack JE, Skow LC, Chowdhary BP: A detailed physical map of the horse Y chromosome. Proc Natl Acad Sci USA 101:9321-9326 (2004b).

Raudsepp T, Chowdhary BP: The horse pseudoautosomal region (PAR): characterization and comparison with the human, chimp and mouse PARs. Cytogenet Genome Res 121:102-109 (2008).

Raudsepp T, Gustafson-Seabury A, Durkin K, Wagner ML, Goh G, Seabury CM, Brinkmeyer-Langford C, Lee EJ, Agarwala R, Stallknecht-Rice E, Schaffer AA, Skow LC, Tozaki T, Yasue H, Penedo MC, Lyons LA, Khazanehdari KA, Binns MM, MacLeod JN, Distl O, Guerin G, Leeb T, Mickelson JR, Chowdhary BP: A 4,103 marker integrated physical and comparative map of the horse genome. Cytogenet Genome Res 122:28-36 (2008a).

Raudsepp T, Paria N, Chowdhary BP: Sequencing-ready physical map of the horse Y chromosome. Poster Abstract P2107: XXXI Conference of the International Society for Animal Genetics (ISAG) Amsterdam, The Netherlands (2008b).

Ravel C, Chantot-Bastaraud S, El Houate B, Berthaut I, Verstraete L, De Larouziere V, Lourenco D, Dumaine A, Antoine JM, Mandelbaum J, Siffroi JP, McElreavey K: Mutations in the protamine 1 gene associated with male infertility. Mol Hum Reprod 13:461-464 (2007).

Reinke V: Sex and the genome. Nat Genet 36:548-549 (2004).

Reymond A, Henrichsen CN, Harewood L, Merla G: Side effects of genome structural changes. Curr Opin Genet Dev 17:381-386 (2007).

Reynard LN, Cocquet J, Burgoyne PS: The multi-copy mouse gene Sycp3-Like Y-Linked (Sly) encodes an abundant spermatid protein that Interacts with a histone acetyltransferase and anacrosomal protein. Biol Reprod (2009) Epub, ahead of print.

Ricchetti M, Tekaia F, Dujon B: Continued colonization of the Human genome by mitochondrial DNA. PLoS Biol 2:1313-1324 (2004).

Rocha D, Affara NA: The genetic basis of impaired spermatogenesis and male infertility. Curr Obst Gyn 10:139–145 (2000).

Rohozinski J, Agoulnik AI, Boettger-Tong HL, Bishop CE: Successful targeting of mouse Y chromosome genes using a site-directed insertion vector. Genesis 32:1-7 (2002).

Rohozinski J, Bishop CE: The mouse juvenile spermatogonial depletion (jsd) phenotype is due to a mutation in the X-derived retrogene, mUtp14b. Proc Natl Acad Sci USA 101:11695-11700 (2004).

Roser JF: Endocrine and paracrine control of sperm production in stallions. Anim Reprod Sci 68:139-151 (2001).

Rosinski KV, Fujii N, Mito JK, Koo KK, Xuereb SM, Sala-Torra O, Gibbs JS, Radich JP, Akatsuka Y, Van den Eynde BJ, Riddell SR, Warren EH: DDX3Y encodes a class I MHC-restricted H-Y antigen that is expressed in leukemic stem cells. Blood 111:4817-4826 (2008).

Rousset H, Chanteloube P, Magistrini M, Palmer E: Assessment of fertility and semen evaluations of stallions. J Reprod Fertil Suppl 35:25-31 (1987).

Rouyer F, Simmler MC, Johnsson C, Vergnaud G, Cooke HJ, Weissenbach J: A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. Nature 319:291-295 (1986).

Salido EC, Yen PH, Koprivnikar K, Yu LC, Shapiro LJ: The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes. Am J Hum Genet 50:303-316 (1992).

Sato Y, Son JH, Tucker RP, Meizel S: The zona pellucida-initiated acrosome reaction: defect due to mutations in the sperm glycine receptor/Cl(-) channel. Dev Biol 227:211-218 (2000).

Saxena R, Brown LG, Hawkins T, Alagappan RK, Skaletsky H, Reeve MP, Reijo R, Rozen S, Dinulos MB, Disteche CM, Page DC: The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. Nat Genet 14:292-299 (1996).

Saxena R, de Vries JW, Repping S, Alagappan RK, Skaletsky H, Brown LG, Ma P, Chen E, Hoovers JM, Page DC: Four DAZ genes in two clusters found in the AZFc region of the human Y chromosome. Genomics 67:256-267 (2000).

Schmahl J, Eicher EM, Washburn LL, Capel B: Sry induces cell proliferation in the mouse gonad. Development 127:65-73 (2000).

Schnieders F, Dork T, Arnemann J, Vogel T, Werner M, Schmidtke J: Testis-specific protein, Y-encoded (TSPY) expression in testicular tissues. Hum Mol Genet 5:1801-1807 (1996).

Schubert S, Dechend F, Skawran B, Krawczak M, Schmidtke J: Molecular evolution of the murine tspy genes. Cytogenet Cell Genet 91:239-242 (2000a).

Schubert S, Dechend F, Skawran B, Kunze B, Winking H, Weile C, Romer I, Hemberger M, Fundele R, Sharma T, Schmidtke J: Silencing of the Y-

chromosomal gene tspy during murine evolution. Mamm Genome 11:288-291 (2000b).

Sekido R, Bar I, Narvaez V, Penny G, Lovell-Badge R: SOX9 is up-regulated by the transient expression of SRY specifically in Sertoli cell precursors. Dev Biol 274:271-279 (2004).

Sekido R, Lovell-Badge R: Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. Nature 453:930-934 (2008).

Sekido R, Lovell-Badge R: Sex determination and SRY: down to a wink and a nudge? Trends Genet 25:19-29 (2009).

Shanker S, Hu Z, Wilkinson MF: Epigenetic regulation and downstream targets of the Rhox5 homeobox gene. Int J Androl 31:462-470 (2008).

Shiue YL, Millon LV, Skow LC, Honeycutt D, Murray JD, Bowling AT: Synteny and regional marker order assignment of 26 type I and microsatellite markers to the horse X- and Y-chromosomes. Chromosome Res 8:45-55 (2000).

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN: A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240-244 (1990).

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC: The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825-837 (2003).

Sun C, Skaletsky H, Birren B, Devon K, Tang Z, Silber S, Oates R, Page DC: An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. Nat Genet 23:429-432 (1999).

Swain A, Narvaez V, Burgoyne P, Camerino G, Lovell-Badge R: Dax1 antagonizes Sry action in mammalian sex determination. Nature 391:761-767 (1998).

Swinburne JE, Boursnell M, Hill G, Pettitt L, Allen T, Chowdhary B, Hasegawa T, Kurosawa M, Leeb T, Mashima S, Mickelson JR, Raudsepp T, Tozaki T, Binns M: Single linkage group per chromosome genetic linkage map for the horse, based on two three-generation, full-sibling, crossbred horse reference families. Genomics 87:1-29 (2006).

Szczerba A, Jankowska A, Andrusiewicz M, Warchol JB: Localization of the DAZ gene expression in seminiferous tubules of patients with spermatogenic disorders. Folia Histochem Cytobiol 44:123-126 (2006).

Szot M, Grigoriev V, Mahadevaiah SK, Ojarikre OA, Toure A, von Glasenapp E, Rattigan A, Turner JM, Elliott DJ, Burgoyne PS: Does Rbmy have a role in sperm development in mice? Cytogenet Genome Res 103:330-336 (2003).

Tanigawa M, Miyamoto K, Kobayashi S, Sato M, Akutsu H, Okabe M, Mekada E, Sakakibara K, Miyado M, Umezawa A, Miyado K: Possible involvement of CD81 in acrosome reaction of sperm in mice. Mol Reprod Dev 75:150-155 (2008).

Tessari A, Salata E, Ferlin A, Bartoloni L, Slongo ML, Foresta C: Characterization of HSFY, a novel AZFb gene on the Y chromosome with a possible role in human spermatogenesis. Mol Hum Reprod 10:253-258 (2004).

Toder R, Glaser B, Schiebel K, Wilcox SA, Rappold G, Graves JA, Schempp W: Genes located in and near the human pseudoautosomal region are located in the X-Y pairing region in dog and sheep. Chromosome Res 5:301-306 (1997).

Torgerson DG, Singh RS: Enhanced adaptive evolution of sperm-expressed genes on the mammalian X chromosome. Heredity 96:39-44 (2006).

Toure A, Grigoriev V, Mahadevaiah SK, Rattigan A, Ojarikre OA, Burgoyne PS: A protein encoded by a member of the multicopy Ssty gene family located on the long arm of the mouse Y chromosome is expressed during sperm development. Genomics 83:140-147 (2004a).

Toure A, Szot M, Mahadevaiah SK, Rattigan A, Ojarikre OA, Burgoyne PS: A new deletion of the mouse Y chromosome long arm associated with the loss of Ssty expression, abnormal sperm development and sterility. Genetics 166:901-912 (2004b).

Toure A, Clemente EJ, Ellis P, Mahadevaiah SK, Ojarikre OA, Ball PA, Reynard L, Loveland KL, Burgoyne PS, Affara NA: Identification of novel Y chromosome encoded transcripts by testis transcriptome analysis of mice with deletions of the Y chromosome long arm. Genome Biol 6:R102 (2005).

Trask BJ, Allen S, Massa H, Fertitta A, Sachs R, van den Engh G, Wu M: Studies of metaphase and interphase chromosomes using fluorescence in situ hybridization. Cold Spring Harb Symp Quant Biol 58:767-775 (1993).

Trask BJ: Human cytogenetics: 46 chromosomes, 46 years and counting. Nat Rev Genet 3:769-778 (2002).

Truong BN, Moses EK, Armes JE, Venter DJ, Baker HW: Searching for candidate genes for male infertility. Asian J Androl 5:137-147 (2003).

Turner ME, Martin C, Martins AS, Dunmire J, Farkas J, Ely DL, Milsted A: Genomic and expression analysis of multiple Sry loci from a single *Rattus norvegicus* Y chromosome. BMC Genet 8:11 (2007).

Van Laere A, Coppieters W, Georges M: Characterization of the bovine pseudoautosomal boundary: Documenting the evolutionary history of mammalian sex chromosomes. Genome Research 18:1884–1895 (2008).

van Roijen HJ, Ooms MP, Spaargaren MC, Baarends WM, Weber RF, Grootegoed JA, Vreeburg JT: Immunoexpression of testis-specific histone 2B in human spermatozoa and testis tissue. Hum Reprod 13:1559-1566 (1998).

Varner DD, Blanchard TL, Brinsko SP, Love CC, Taylor TS, Johnson L: Techniques for evaluating selected reproductive disorders of stallions. Anim Reprod Sci 60-61:493-509 (2000).

Varner DD, Johnson L: From a sperm's eye view - revisiting our perception of this intriguing cell. AAEP Proceedings 53:104-177 (2007).

Vicdan A, Vicdan K, Gunalp S, Kence A, Akarsu C, Isik AZ, Sozen E: Genetic aspects of human male infertility: the frequency of chromosomal abnormalities and Y chromosome microdeletions in severe male factor infertility. Eur J Obstet Gynecol Reprod Biol 117:49-54 (2004).

Vodicka R, Vrtel R, Dusek L, Singh AR, Krizova K, Svacinova V, Horinova V, Dostal J, Oborna I, Brezinova J, Sobek A, Santavy J: TSPY gene copy number as a potential new risk factor for male infertility. Reprod Biomed Online 14:579-587 (2007).

Vogel T, Borgmann S, Dechend F, Hecht W, Schmidtke J: Conserved Y-chromosomal location of TSPY in bovidae. Chromosome Res 5:182-185 (1997a).

Vogel T, Dechend F, Manz E, Jung C, Jakubiczka S, Fehr S, Schmidtke J, Schnieders F: Organization and expression of bovine TSPY. Mamm Genome 8:491-496 (1997b).

Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Kohn FM, Schill WB, Farah S, Ramos C, Hartmann M, Hartschuh W, Meschede D, Behre HM, Castel A, Nieschlag E, Weidner W, Grone HJ, Jung A, Engel W, Haidl G: Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. Hum Mol Genet 5:933-943 (1996).

Vogt PH, Falcao CL, Hanstein R, Zimmer J: The AZF proteins. Int J Androl 31:383-394 (2008).

Vong QP, Li Y, Lau YF, Dym M, Rennert OM, Chan WY: Structural characterization and expression studies of Dby and its homologs in the mouse. J Androl 27:653-661 (2006).

Wade CM, Gnerre S, Garber M, Lander E, Lindblad-Toh K: Preliminary analysis of the equine genome sequence: 7th Dorothy Russell Havemeyer International Equine Genome Mapping Workshop, Abstarct p6, Tahoe City, California (2007).

Wallner B, Piumi F, Brem G, Muller M, Achmann R: Isolation of Y chromosome-specific microsatellites in the horse and cross-species amplification in the genus Equus. J Hered 95:158-164 (2004).

Wang PJ: X chromosomes, retrogenes and their role in male reproduction. Trends Endocrinol Metab 15:79-83 (2004).

Waters PD, Ruiz-Herrera A, Dobigny G, Garcia Caldes M, Robinson TJ: Sex chromosomes of basal placental mammals. Chromosoma 116:511-518 (2007).

Wilhelm D, Martinson F, Bradford S, Wilson MJ, Combes AN, Beverdam A, Bowles J, Mizusaki H, Koopman P: Sertoli cell differentiation is induced both cell-autonomously and through prostaglandin signaling during mammalian sex determination. Dev Biol 287:111-124 (2005).

Wilhelm D: R-spondin1--discovery of the long-missing, mammalian female-determining gene? Bioessays 29:314-318 (2007).

Wilhelm D, Hiramatsu R, Mizusaki H, Widjaja L, Combes AN, Kanai Y, Koopman P: SOX9 regulates prostaglandin D synthase gene transcription *in vivo* to ensure testis development. J Biol Chem 282:10553-10560 (2007a).

Wilhelm D, Palmer S, Koopman P: Sex determination and gonadal development in mammals. Physiol Rev 87:1-28 (2007b).

Wong EY, Tse JY, Yao KM, Tam PC, Yeung WS: VCY2 protein interacts with the HECT domain of ubiquitin-protein ligase E3A. Biochem Biophys Res Commun 296:1104-1111 (2002).

Woods J, Rigby S, Brinsko S, Stephens R, Varner D, Blanchard T: Effect of intrauterine treatment with prostaglandin E2 prior to insemination of mares in the uterine horn or body. Theriogenology 53:1827-1836 (2000).

Yen PH, Chai NN, Salido EC: The human DAZ genes, a putative male infertility factor on the Y chromosome, are highly polymorphic in the DAZ repeat regions. Mamm Genome 8:756-759 (1997).

Young AC, Kirkness EF, Breen M: Tackling the characterization of canine chromosomal breakpoints with an integrated in-situ/in-silico approach: the canine PAR and PAB. Chromosome Res 16:1193-1202 (2008).

Yu YE, Zhang Y, Unni E, Shirley CR, Deng JM, Russell LD, Weil MM, Behringer RR, Meistrich ML: Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. Proc Natl Acad Sci USA 97:4683-4688 (2000).

Zechner U, Wilda M, Kehrer-Sawatzki H, Vogel W, Fundele R, Hameister H: A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution? Trends Genet 17:697-701 (2001).

Zhang P, Compagnone NA, Fiore C, Vigne JL, Culp P, Musci TJ, Mellon SH: Developmental gonadal expression of the transcription factor SET and its target gene P450c17 (17 alpha-hydroxylase/c17, 20 lyase). . DNA Cell Biol 20:613-624 (2001).

Zhao L, Burkin HR, Shi X, Li L, Reim K, Miller DJ: Complexin I is required for mammalian sperm acrosomal exocytosis. Dev Biol 309:236-244 (2007).

Zhao L, Reim K, Miller DJ: Complexin-I-deficient sperm are subfertile due to a defect in zona pellucida penetration. Reproduction 136:323-334 (2008).

WEBSITE REFERENCES

1. http://www.informatics.jax.org/; Mouse Genome Informatics (MGI) Homepage.

2. http://www.ncbi.nlm.nih.gov/; National Center for Biotechnology Information (NCBI) Homepage.

3. http://www.ensembl.org/index.html; Ensembl Genome Browser Homepage.

4. http://genome.ucsc.edu/; UCSC Genome Bioinformatics Homepage.

5. http://www.ncbi.nlm.nih.gov/mapview/; NCBI Map Viewer Homepage.

6. http://www.horsecouncil.org; American Horse Council Homepage.

7. http://www.broad.mit.edu/mammals/horse/; Broad Institute Horse Genome Project Homepage.

8. http://www.repeatmasker.org; Repeat Masker Homepage.

9. http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi; NCBI BLAST Homepage.

10. http://www.ensembl.org/Equus_caballus/index.html; Ensembl Genome Browser Horse Genome Assembly Homepage.

11. http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi; Primer 3 Homepage.

12. http://www.ncbi.nlm.nih.gov/projects/gorf/; NCBI ORF Finder Homepage.

13. http://bacpac.chori.org/equine241.htm; Children's Hospital Oakland Research Institute (CHORI) 241 Equine BAC Library Homepage.

14. http://hbz7.tamu.edu/homelinks/bac_est/bac.htm; Texas A&M University BAC Library Homepage.

15. http://www.genome.wustl.edu/tools/?overgo.html; Genome Sequencing Center Homepage.

16. http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide; NCBI Entrez Nucleotide Homepage.

17. http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi; NCBI Conserved Domains Homepage.

APPENDIX

Table A1: Detailed information about horse MSY BAC end sequence (BES) analysis.

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
002E4-SP6	CT864643	701	32.4			LINE1
002E4-T7	CT864644	714	36.1		(AC)16, (AT)11	LTR/ERV1, simple repeat, SINE
003.4H8-M13		880	46.1			no repeats
003.4H8-T7		979	43.9			LINE1
003A5-SP6	CT825159	631	39.1			SINE
003A5-T7	CT825160	706	43.2			MER34C, LTR/ERV1
005.2A8-M13		889	39.1			DNA/MER1_type, DNA/AcHobo, LINE1
005.2A8-T7		878	45.9			LTR/ERVL
006B10-SP6	CT827035	387	43.7			no repeats
006B10-T7	CT826936	1031	38.3			DNA/MER2
008.3G9-M13		632	41.8			no repeats
008.3G9-T7		758	39.1			SINE
011B8-SP6	CT840125	749	37.9		(CT)3 CC (CT)1 CC (CT)7	Simple repeat, LTR/MaLR
011B8-T7	CT840126	769	36.4			no repeats
012.2E5-M13		716	31.4			DNA/MER2
012.2E5-T7		281	42.3			SINE
012G3-SP6	CT841047	282	38.7			simple repeat, LTR/MaLR
012G3-T7	CT841048	456	49.8			no repeats
013E2-SP6	CT865642	1323	39.2			LINE1
013E2-T7	CT865502	1313	34			LINE1, LTR/ERV1
015.2A9-M13		635	44.9			LTR/ERVL
015.2A9-T7		535	33.3			low complexity
015.2E9-M13		445	41.8			SINE
015.2E9-T7		389	34.7			LINE1

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
016.4C5-M13		182	52.7			no repeats
016.4C5-T7		312	35.9			LINE1
017.2C11-M13		869	37.1			LINE1
017.2C11-T7		712	31.2			LINE1, low complexity repeat
017D15-SP6	CT939410	636	47.5			LINE1
017D15-T7	CT939412	742	33.8			LINE2
018K2-SP6		865	36.3			no repeats
018K2-T7		855	37.2			LINE1
019D21-SP6	CT961235	882	42.1			LTR/MaLR, SINe/MIR, LINE1
019D21-T7	CT961236	524	46			LTR/ERVL, LINE2
020.1G12-M13		858	43.7			Low complexity repeat
020.1G12-T7		777	45.2			no repeats
020B4-SP6	CT938442	502	39.8			no repeats
020B4-T7	CT938444	342	33.6			LINE1
020L18-SP6	CT939014	756	39.6			LTR/ERV1
020L18-T7	CT939015	567	44.4		(TG)7 CG (TG)7	LINE1, simple repeat, SINE
022.4 E3-M13		822	47.9			LINE1
022.4 E3-T7		706	35.8			LINE1
022G3-SP6	CT941801, ET052928	400	47.3			no repeats
022G3-T7	CT941802, ET052929	720	53.7			no repeats
022P7-SP6	CT942195	744	52			no repeats
022P7-T7	CT942196	468	32.9			LINE1, DNA/MER2
024.4G8-M13		760	49.7			SINE, LINE1
024.4G8-T7		939	37.2			no repeats

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
024I23-SP6	CT942381	849	47.9			no repeats; but amplifies both male and female
024I23-T7	CT942288	928	38.4			no repeats
026.4A6-M13		311	35.4			LINE1
026.4A6-T7		347	45.8			no repeats
026B21-SP6	CT951493	714	43.4			LINE1
026B21-T7	CT951494	783	38.7			LINE1
027.1A2-M13		1169	40.5			LINE1
027.1A2-T7		294	41.2			no repeats
027A12-SP6	CT952155	772	38.7			no repeats
027A12-T7	CT952156	761	32.5			LTR/MaLR, simple repeat
027B13-SP6		698	42.4			LTR/MaLR
027B13-T7	CT952199	555	38.7			no repeats
032H24-SP6	CT957174	638	39.5			LINE2
032H24-T7	CT957175	444	49.8			SINE/MIR
032K15-SP6	CT957292	633	47.1			LINE2, simple repeat
032K15-T7	CT957293	693	41			no repeats
034 E15-SP6	CT958457	605	39.8			no repeats
034 E15-T7	CT958458	588	44.5			no repeats
037.4D11-M13		558	39.8			LINE1
037.4D11-T7		637	44.4			LINE1
037D12-SP6	CT960488	814	36.2			LINE1
037D12-T7	CT960489	667	31.8			LINE1
039P6-SP6		334	43.7			no repeats
039P6-T7		619	40.7			no repeats
041O19-SP6	CT967953	682	34.9			LINE1
041O19-T7	CT967954	508	31.9			low complexity
042.4B5-M13		785	48.7			LINE1
042.4B5-T7		968	34.3			LINE1

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
047.1H3-M13		332	36.4		(TG)14	LINE1, simple repeat
047.1H3-T7		704	35.8			LTR/ERV1
047.2A3-M13		701	40.8			no repeats
047.2A3-T7		855	36.1			LTR/ERVL, SINE, LINE1, DNA/MER2
047B7-SP6	CT964598	422	36.3			LINE1, DNA/hAT-Charlie
047B7-T7	CT964599	384	39.6			LINE1
049.2F10-M13		1016	41.6			SINE. MIR
049.2F10-T7		1181	45			SINE
049.3F11-M13		955	35.4			SINE, ERE1
049.3F11-T7		884	38.8	CRISP3, 2e-06		no repeats
049J16-SP6	CT969091	753	32			LINE1
049J16-T7	CT969092	641	35.7			LINE1
052H5-SP6	CT972580	684	35.1			LINE1, low complexity
052H5-T7	CT972585	573	38.6			no repeats
054A8-SP6	CT973016	462	42.2			LINE1
054A8-T7	CT973017	455	44.4			no repeats
054F13-SP6	CT973230	703	43			no repeats
054F13-T7	CT973231	542	33.9			LTR/ ERV1
054J7-SP6	CT973393	670	35.1			LINE1
054J7-T7	CT973394	620	50.2			no repeats
055N19-SP6		720	30.4			LINE1
055N19-T7		800	36.7			LINE1
060D8-SP6	CT976605	856	39			SINE, LINE1
060D8-T7	CT976606	645	44.3			LTR element 168 bp in the end
061.4H6-M13		931	41.6			simple repeats
061.4H6-T7		940	38.4			LINE1, SINE
061G21-SP6	CT977473	335	62.7			no repeats
061G21-T7	CT977474	368	31			SINE

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
061J18-SP6	CT977591	201	37.3			no repeats
061J18-T7	CT977592	395	34.4			LINE1
063H12-SP6	CU002478	253	45			no repeats
063H12-T7	CU002479	803	35.7			LINE1
063I4-SP6		353	47.9			SINE
063I4-T7	CU002511	922	42.6			SINE
064P16-SP6	CU003552	685	35.2			LINE1
064P16-T7	CU003553	610	64.4			no repeats
066M24-SP6	CU001288	412	40.3			no repeats
066M24-T7	CU001289	511	46.8			no repeats
067.1G8-M13		712	35	NLGN4Y 85.0% NLGN4X 86.6% (BLAT)		no repeats
067.1G8-T7		716	28.9			simple repeat, low complexity, LINE1
067.4G1-M13		1185	43.4			SINE, LINE1
067.4G1-T7		Sequence not available				
069 E11-SP6	CU000182	757	50.2			LINE1
069 E11-T7		479	47.6			no repeats
070F17-SP6	CU138007	264	50		(CTCA)5	LTR/MaLR, simple repeat
070F17-T7	CU138008	541	42			NCBILTR/ERV1, SINE
072G23-SP6	CU004599	824	37.9			LINE1
072G23-T7	CU004600	765	35.6			LINE1
072G7-SP6	CU004572	783	36.4			LINE1
072G7-T7	CU004573	764	39.1			no repeats
074P12-SP6	CU006411	780	42.2			LINE1
074P12-T7	CU006412	439	33.6			LINE1
077M19-SP6	CU006998	596	40.3			LINE1, LTR, low complexity
077M19-T7	CU006999	305	44.3			no repeats

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
079.4H1-M13		718	41.4			no repeats
079.4H1-T7		621	35.6			LINE1
080.4F7-M13		499	45.5			SINE, LINE1
080.4F7-T7		544	41.4			no repeats
081F24-SP6	CU011589	325	41.2			LINE1
081F24-T7	CU011590	465	36.8			LINE1
081F8-SP6	CU011557	804	38.1			no repeats
081F8-T7	CU011558	648	36.7			no repeats
083H5-SP6	CU017259	701	32.1			LINE1, low complexity
083H5-T7	CU017260	751	32.8			LINE1, DNA/MER1, low complexity, simple repeat
086.2F8-M13		1002	37.4			SINE
086.2F8-T7		1088	39.2			SINE, DNA/MER1, LTR/MaLR
086J1-SP6	CU022552	667	29.8			LINE1, LTR
086J1-T7	CU022553	575	34.4			LINE1
087.3A5-M13		867	40.9			no repeats
087.3A5-T7		1057	37			LINE1
089.3B11-M13		483	39.5			LTR and LINE1
089.3B11-T7		833	39.7			DNA/MER1, DNA/AcHobo, LINE1
090B11-SP6	CU015692	492	31.5			LINE1
090B11-T7	CU015693	396	39.7			LTR/ERVL
090G18-SP6	CU015913	737	42.2			LINE1
090G18-T7	CU015914	691	41.8			LTR/MaLR, LINE1
090P8-SP6	CU016271	510	45.3			SINE, LTR
090P8-T7	CU016272	524	35.1			LINE1
091.4G10-M13		568	40.3			LTR/ERV1

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
091.4G10-T7		782	39.3			LINE1
095.4B8-SP6		sequence not available	n/a			n/a
095.4B8-T7		193	36.3			no repeats
095.4F10-M13		690	47			LTR element
095.4F10-T7		607	44			no repeats
097D2-SP6	CU019809	854	38.9		(AT)2 C (AT)25	simple repeat, SINE, DNA/MER
097D2-T7	CU019810	821	37.8	CRISP1 3e-75		SINE
100.3A11-M13		644	44.3			no repeats
100.3A11-T7		718	37.2			LINE2
100.4F5-M13		681	41.3			LINE1
100.4F5-T7		716	36.5			LINE2
100H13-SP6	AJ542956	868	43.2			LTR/MaLR
100H13-T7	AJ542957	804	34.1			LINE1
101H8-SP6	AJ576754	1032	37.7			SINE
101H8-T7	AJ576755	1164	39.6			LINE1
102J15-SP6	AJ584337	837	33.2			low complexity
102J15-T7	AJ584338	882	37.8			LINE1
103.3A6-M13		1078	46.2			SINE, LINE1
103.3A6-T7		sequence not available	n/a			n/a
106F1-SP6	CT007963	899	46.9			LINE1
106F1-T7	CT008006	873	40.8			LINE1
106J17-SP6		743	44.4	NLGN4X 0.0		no repeats
106J17-T7		772	39.5			LINE/CR1
107.3H9-M13		816	39.3			LINE1, SINE
107.3H9-T7		763	52.6			LINE1
108.4C7-M13		490	50.4			no repeats

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
108.4C7-T7		1150	44			no repeats
110.3H12-M13		1042	37.2	EIF2S3 1e-63		LTR/MaLR, snRNA
110.3H12-T7		1002	40			simple repeat
111.2F5-M13		1044	41.9			LINE1, SINE
111.2F5-T7		1182	45			LINE1
112.1A9-M13		1443	45.6			LINE1
112.1A9-T7		1450	49.4			simple repeat
112C10-SP6		585	37.8			LINE1
112C10-T7		913	38.2			LINE1
112E12-SP6	CR955686	587	49.7			no repeats
112E12-T7	CR957122	494	36.4			LTR/MaLR
114E24-SP6	CT006488	636	37.4			LINE1
114E24-T7	CT006811	720	38.1			DNA/MER1
114I17-SP6	CT006994	431	34.6			SINE/MIR
114I17-T7	CT007019	463	40.2			no repeats
117.4F7-M13		928	32.5			MER2
117.4F7-T7		757	38.6			SINE and LINE1
118.1A9-M13		781	41.6			LINE2
118.1A9-T7		680	28.1			LINE1
118L7-SP6	CU025610	840	47.7			LINE1
118L7-T7	CU025611	1111	37.8		(AC)5 AT (AC)13	simple repeat
118N21-SP6	CU025698	396	40.4			no repeats
118N21-T7	CU025699	745	34.5			no repeats
119K22-SP6	CU026858	314	27.7			LINE2
119K22-T7		426	36.6			LINE1, SINE/Alu
120.1A5-M13		1042	34.8	AMELY		DNA/tip100
120.1A5-T7		648	38.5			SINE, LINE1
120A19-SP6	CU027080	220	55			LINE1

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
120A19-T7	CU027081	362	58.3			no repeats
121G24-SP6	CU028694	622	42.8			no repeats
121G24-T7	CU028695	523	33.3			LINE1
121H9-SP6	CU028712	513	31.2			LINE1, SINE
121H9-T7	CU028713	793	34.4	horse MAP3K7IP3 8e-76; ECAX 24.1. Human MAP3K7IP3 1e-66; HSAX 30.7;		no repeats
124.3G9-M13		477	33.5			DNA/MER2
124.3G9-T7		852	39.4			LTR/MaLR
125.3G11-M13		711	42.3			SINE, LTR, MER2
125.3G11-T7		644	39			no repeats
125H6-SP6	CU029932	674	26.3			LINE1
125H6-T7		847	40.3			LINE1, SINE
126G2-SP6	CU029327	522	39.7			no repeats
126G2-T7		866	37.6			SINE, LTR present
129K23-SP6	CU033122	528	53.6			no repeats
129K23-T7	CU033123	409	46			LINE1
131N23-SP6	CU034701	559	33.8			SINE, LINE1
131N23-T7	CU034702	590	29			LINE1
132K10-SP6	CU035255	560	42			DNA/MER1, SINE
132K10-T7	CU035256	590	38.8			LTR/ERV1
132N15-SP6	CU035402	533	41.8			no repeats
132N15-T7	CU035403	551	43.8			no repeats
134H14-SP6	CU036544	623	62.9			no repeats
134H14-T7	CU036545	584	37.2			LINE1
134I16-SP6	CU036592	688	55.2			no repeats
134I16-T7	CU036593	667	49			no repeats

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
137I17-SP6	CU038728	423	44.9			no repeats
137I17-T7	CU038729	484	45.7			DNA/MER2
139C20-SP6	CU040631	686	46.8			LINE1
139C20-T7	CU040632	746	41.8			LINE1
140J20-SP6	CU039506	801	40.6			LINE1
140J20-T7	CU039507	817	38.8			LINE1
140M23-SP6	CU039648	885	36.8			LINE1
140M23-T7	CU039649	879	30.8			LINE1
142O2-SP6	CU044834	569	40.1			LINE1
142O2-T7	CLY074, CU044835	529	50.3			no repeats
144B9-SP6	CU045683	711	40.4			LINE1
144B9-T7	CU045684	508	49.4			no repeats
145I6-SP6	CU046684	781	35.1			LINE1
145I6-T7	CU046685	624	41.7			LINE1
147K8-SP6	CU048189	792	39.8			LINE1
147K8-T7	CU048190	770	36.9			LINE1
148G3-SP6	CU048698	509	36.1			LINE1
148G3-T7	CU048699	1218	48.7			low complexity
149H18-SP6	CU049495	633	54.3			no repeats
149H18-T7	CU049496	652	36.5			LINE1
152E2-SP6		772	35.6			no repeats
152E2-T7		429	42.2			LTR/MaLR
152G20-SP6	CU051535	569	54			LINE1
152G20-T7	CU051536	590	37.5			LINE1
155B8-SP6		851	65.7			no repeats
155B8-T7		949	37.7			LINE/RTE
155M11-SP6	CU054469	604	41.7			LINE1

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
155M11-T7	CU054470	673	38.3			LTR/MaLR, SINE/MIR
159E3-SP6	CU056230	551	36.5			no repeats
159E3-T7	CU056231	555	27			SINE
159F5-SP6	CU056277	591	27.1			no repeats
159F5-T7	CU056278	576	32.1			DNA/MER1
160K10-SP6	CU057249	719	28.9			low complexilty
160K10-T7	CU057250	758	60.4			LINE2
165 E24-SP6	CU060545	660	45.6			no repeats
165 E24-T7	CU060546	580	37.4			LINE1
167N20-SP6	CU062306	385	34.8			Low complexity, SINE/MIR
167N20-T7	CU062307	630	31.7			low complexity
168I4-SP6	CU061376	929	35.8			LINE1
168I4-T7		1008	39.8			LTR/ERVL, LINE1
168O8-SP6	CU061645	609	49.1			LINE1
168O8-T7	CU061646	551	37.1			LTR/ERVL-MaLR, LINE1
172 E14-SP6	CU079797	549	39			no repeats
172 E14-T7	CU079798	626	35.1			LINE1, low complexity
172D14-SP6	ET052930	354	56.2			no repeats
172D14-T7	ET052931	539	35.1			LINE1, low complexity
172I8-SP6		657	38.5			LINE1
172I8-T7		839	38.9			no repeats
179K8-SP6	CU083108	560	61.4			low complexity, LINE1, Alu
179K8-T7	CU083388	604	48.3			LTR/ERVL, SINE
180P20-SP6	CU084740	777	36.3			DNA/MER1, low complexity
180P20-T7	CU084545	780	35.8			LTR/ERV1
181B18-SP6	CU085205	506	43.7			SINE, LTR
181B18-T7	CU085127	567	44.4			LINE1
185M14-SP6	CU087822	684	34.9			no repeats

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
185M14-T7	CU088298	390	36.4			LINE1
186J13-SP6		320	39			LINE1
186J13-T7		560	36			LINE1
188E20-SP6	CU088810	693	41.4			SINE
188E20-T7	CU089040	1131	48.2			LINE1, low complexity
190M2-SP6	CU091407	1150	40.4			LINE1
190M2-T7	CU091759	1124	37.7			no repeats
205D10-SP6	CU100955	552	41.5			LINE1, LINE2
205D10-T7	CU100417	522	35.6	CRISP3/TPX1 6e-12		LINE1
207D10-SP6	CU102230	660	38.2			LINE1
207D10-T7	CU102329	649	61.8			no repeats
207P5-SP6	CU102452	632	51.1			no repeats
207P5-T7	CU102185	649	46.1			LTR, ERV1
209K10-SP6	CU103461	777	34.7	KAL1 4e-112		no repeats
209K10-T7	CU103487	761	34.4			LINE1
215C6-SP6	CU108088	632	36.6			SINE/MIR
215C6-T7	CU107955	605	45.1			DNA/Tip100, SINE
244B13-SP6	CU126429	525	44.4			DNA/MER2
244B13-T7	CU126281	306	50.3			no repeats
263G23-SP6	CU140369	698	47.9			simple repeat
263G23-T7	CU139968	632	46.4			no repeats
264G20-SP6	CU141122	658	60.3			Low_complexity
264G20-T7	CU143373	612	44.4			DNA/TcMar-Tigger
269J9-SP6	CU143890	758	30.6			SINE/MIR
269J9-T7	CU143890	629	39.9			LTR/ERVL, DNA/MER1
272B4-SP6	CU147224	611	33.2			LINE1
272B4-T7		872	32.8			LINE1

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
275P16-SP6	CU148527	799	33.4			no repeats
275P16-T7	CU148738	760	38.9			LINE1
278M12-SP6	CU148483	593	33.4	UTY (BLAT)		no repeats
278M12-T7	CU148207	660	52			no repeats
280P20-SP6	CU151385	349	55.6			LINE1
280P20-T7	CU151203	655	33.9			LINE1
291D19-SP6	CU152144	645	32.4			low complexity, LINE1
291D19-T7	CU152471	656	35.7			LINE1
309A2-SP6	CU168483	528	47			no repeats
309A2-T7	CU168709	482	48.1			no repeats
318M1-SP6	CU168992	584	31.8			LTR/ERV1
318M1-T7	CU169153	558	37.3			no repeats
324H11-SP6	CU207515	627	55.7			no repeats
324H11-T7	CU208053	566	49.5			LINE1
325B20-SP6	CU208562	682	45.2			DNA/hAT-Charlie, LTR/ERVL-MaLR
325B20-T7	CU208331	687	48.5			SINE/MIR
329G16-SP6	CU210328	678	31.7			MER1
329G16-T7	CU210572	670	34			MER2
331E10-SP6	CU199001	697	40.6			LTR/MaLR
331E10-T7	CU199033	716	38.3	CRISP3/TPX1 8E-39		LTR/ERV1, LTR/MaLR
335P13-SP6	CU184042	708	34.6			SINE/MIR
335P13-T7	CU184277	684	33			LINE1
338A6-SP6	CU203349	648	27.2			no repeats
338A6-T7	CU203043	673	30.6			no repeats
341G20-SP6	CU204940	648	36.1			LINE1
341G20-T7	CU204493	678	33.3			LINE1
344A12-SP6	CU312820	767	37.3			no repeats

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
344A12-T7	CU312716	738	36.6			LINE1
377O23-SP6	CU236334	653	36			no repeats
377O23-T7	CU236809	581	42			LTR/ERVL
394K12-SP6	CU246464	663	43.1			LINE1
394K12-T7	CU246587	621	48.1			no repeats
395L19-SP6	CU248760	594	44.9			no repeats
395L19-T7	CU248652	526	33.5			LINE1
406I22-SP6	CU257716	625	36.3			LINE1
406I22-T7	CU258117	483	40.2			LINE1
415H8 -SP6	CU264902	687	37.1			LINE
415H8 -T7	CU264886	577	38			LINE1
417N24-SP6	CU265590	674	35.2			SINE/MIR
417N24-T7	CU265575	725	35.3			no repeats
418J18-SP6	CU267189	750	46.9			no repeats
418J18-T7	CU266817	754	47.9			LINE1
422E23-SP6	CU271235	598	38.1			no repeats
422E23-T7	CU271261	567	32.3	KAL1 4e-97		no repeats
437I11-SP6	CU301860	716	34.8			SINE/MIR
437I11-T7	CU301866	679	33.9			SINE
450C22-SP6	CU286553	334	33.8			LINE1
450C22-T7	CU286365	314	33.8			no repeats
456J9-SP6	CU289788	683	41.4			LINE1, SINE
456J9-T7	CU289914	731	38.6			LINE1
504H13-SP6		856	31.1	ZFY 4e-29		low complexity repeats
504H13-T7		869	33.8			LINE1
510F11-SP6		854	32.7			no repeats
510F11-T7		788	36.4			LINE1
ABW-M13		193	52.3			simple repeats
ABW-T7		209	38			no repeats

Table A1 continued

BES	GenBank Acc. No.	Size, bp	GC%	BLAST significant similarity (<E-10) to known genes	Microsatellites	Other repeats
BBW-M13		157	43.9			no repeats
BBW-T7		214	35.5			LINE1
CBW-M13		177	46.3			no repeats
CBW-T7		768	34.1			LINE1
DBW-M13		150	52			LINE1
DBW-T7		195	50.8			simple repeats
EBW-M13		693	41.4			no repeats
EBW-T7		1131	48.2			no repeats
FBW-M13		1056	39.7			DNA/MER2
FBW-T7		1385	42.2			LINE1, LTR/ERV1
GBW-M13		993	39.6			LINE1, SINE
GBW-T7		990	37.5			no repeats
HBW-M13		247	52.2			simole repeats, LTR
HBW-T7		232	47.8			no repeats
JBW-M13		238	55			no repeats
JBW-T7		660	39.3			LINE1
LBW-M13		642	34.6			LINE1
LBW-T7		1121	39			LTR
MBW-M13		1146	43.5			SINE
MBW-T7		1092	37.5			LINE1
OBW-M13		1055	43.3			LINE1
OBW-T7		1094	44.1			no repeats

Table A2: Detailed information about the STS markers on horse MSY contig map.

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
002E4-T7	F: CCACAGCAGTGATGAGATGG R: TGCCAAGTTTTACCTGGTCTC	218	58	M/wF	
003.4H8-M13	F: GATTAAGAGCCCAGGAGAGG R: TTGGCTGGAGACTTGGTAGG	250	64	M	
003.4H8-T7	F: GCAACAGAAAGCACAGAGG R: TGCCAGGACCATCTCAGG	174	62	M	
003A5INT1	F: TCAGTGGGTCTGTTTCATGC R: TTGCCTCCATGATTTTCTCC	157	58	M/wF	
003A5INT2	F: GCTACAGCTTGGTCCTCTGC R: AGTGCAATACGGGAGTCAGG	241	58	M/wF	
003A5-SP6	F: GCCCCATCCATCAGTTTTTA R: TGCATTTCTTCATTCCACTCC	157	58	M	CT825159
005.2A8-M13	F: TCAGAATGGAGCTGCCTAAAA R: GGCCCTGGCTTTTCTTTATT	274	58	M	
005.2A8-T7	F: GTCAATCCTGCTGCCCTTAG R: TGAGCAAGCAAATGGAAAGA	247	58	M	
006B10-SP6	F: GCGAGGTGGCTTCTCTTATG R: TACGCAGTTGTCGAACTTGG	199	58	M	CT827035
006B10-T7	F: CTTGCAATCACGTGGAAGAA R: TTGAATGCCACAGGTAAGGA	207	TD60	M	CT826936
008.3G9-M13	F: TTGTAGGCATTGTGCCAGTT R: GGGCTTGTAAGAGACCCACA	243	58	M	
008.3G9-T7	F: TGACAACATTCTGGCAGGAG R: TGAGCACCCAAACCATGATA	161	58	M/wF	
011B8-SP6	F: GATGATAAGTGCTCTTCATTTGTGA R: TGGGAAAAGTGGTTGGAGTG	232	58	M	CT840125

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
011B8-T7	F: ACCGGCTTGAGAGAGAATGA R: ACCTTTGGGAAAATGCCTCT	174	58	M	CT840126
012.2E5-M13	F: TGGGTTTCTTGGAATCCTGA R: CCAGTGAAGGGGGACATCA	249	60	M	BV140783, CLY041
012.2E5-T7	F: CCCCTCTTGCTGAGTTTTTG R: ACAGCAAACCAAACCCTACG	150	60	M	BV140831, CLY037
012G3-SP6	F: AGCAGCCTTCTAGCTTCGTT R: CTTGTGCCCTCCATTTTTGT	204	58	M/wF	CT841047
012G3-T7	F: CCATCCAAATCTGTCCTGCT R: CCCAGCAGACCTTGTTTGTT	213	58	M/F	CT841048
013E2-T7	F: CAGACCAGAAGCTGAAGAAGAG R: GGGCTGCATACAAGGAAAGT	331	65	M	BV140783, CLY042
015.2A9-M13	F: CCCATGACCTGTCCATACTG R: AACCAAGCCACATTTCATCG	182	60	M	BV140824, CLY028
015.2A9-T7	F: GGAGGCCACAGAGTGTTTTT R: GAAAGGTTGTCTCCATCTTTCCT	185	60	M	BV140822, CLY022
015.2E9-M13	F: GCCAGTACATGGCCAGAGTT R: GGAGCTCTGTGAATGGAAGC	159	50	M	
016.4C5-M13	F: GTGGCAGTCTGGGTTACGTT R: AAGCAAACCAAGGAGAAGCA	215	50	M	
017.2C11-M13	F: CCCAAAGTGGAATGTGAGAG R: AAATTACCACTTGTGTAAGGTGAACA	150	58	M/F	
017.2C11-T7	F: TCATCATTCTTTGGCATTTG R: CCAACACTCAGGCAATTTGA	194	58	M	

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
017D15-SP6	F: GCTGCCTGAAACCTGACTTC R: TACTCCAAGTGCCAAGCTCA	176	65	M	BV140807, CLY083
017D15-T7	F: AGTGGGTTTCAGGCAAAATG R: GGCACTTGTCTTCTCGGTGT	209	65	M	BV140806, CLY082
018K2-SP6	F: GGTTTGACCCAAGCAAAGAA R: TGGGAAATGAAATTGCACAC	209	58	M/F	
019D21-SP6	F: AGCTCCTCTGGGCACCTATC R: GGAAAAACTGCTCCATTCCTC	111	58	M	
019D21-T7	F: TTTTCCTTGGCCTTTACTCC R: GCAAAGAATTTAGGCCTGGT	217	60	M	
020.1G12-T7	F: TCTGGGTTCTGGATCTGACTG R: GACTCGGCCTGAAGCTAATG	173	62	M	
020L18-SP6	F: TCAATAGCCATGGTGAGCTG R: TTCCAACCTCATTCCCTTTG	166	58	M	
020L18-T7	F: GCTCCTTCTTGTGGCACAGT R: TCTTTGTAATCAGTAGCCCCATT	407	58	M	
022.4 E3-M13	F: CCGATTCCAAACCATGAGAT R:AGGAAGTCAGCACCTTGCAT	138	65	M	BV140843, CLY066
022G3-SP6	F: ACCTCAGTAGGGGGCTTCTC R: GTGGAACAGGGTAAGGCAAA	154	58	M	
022P7-SP6	F: GCAATGGGACTGTGGAAAAT R:TTTTCACCTCAGCCCTCAAC	383	151	M/F	
024.4G8-M13	F: ACAGTCTCCTGCTGGTTTCC R: AATTTCAGCCTCCTTTCCA	156	60	M	BV140847, CLY091
024.4G8-T7	F: TCATCTCGTATCTCCTCATATCC R: TCCCTATCCTTGTTGAAAATCC	495	60	M	BV140846, CLY090

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
024I23-T7	F: ATCTGCTTCGGCCTTCTCTT R: GGTGTACCCTGCTTCTCGT	101	65	M	BV140808, CLY101
026.4A6-M13	F: CTCCCCTCCTCCCACATTAT R: GGCAGCAGATCAACTACCTG	153	60	M	BV140825, CLY029
026.4A6-T7	F: TCGGTGACATCAGCAAAATG R: TACCCCGAATCCAGATCCTC	192	55	M	BV140823, CLY026
026B21-SP6	F: CAAGGAAGCCAGGAAGAGTG R: GTCTCTGGCCCATGAGTGAT	160	60	M	
027.1A2-M13	F: AACACCCAACCACATAGAGGA R: GGGCCGTTGTTGAGTCTTAG	190	60	M	
027.1A2-T7	F: AGGGAGGTCTATGGAGAAGG R: ATTTTTAGCTTGCCCTTTGG	296	65	M	BV140841, CLY059
027A12-SP6	F: GCCCATGCTAAATTTGTGCT R: CTAATTGAGGGGCAACCAAA	237	58	M	CT952155
027A12-T7	F: CATTTTTGCTGTTTCCCACA R: CAGGAAGACAACCAGAAAAACC	162	58	M	CT952156
027B13-SP6	F: GAAACTGCCAGTGAGACAAGG R: TGTCTGGCATAGGAACTCAAG	698	58	M/wF	
027B13-T7	F: CATGCTGTAATAAGACTGAGAAGA R: ACAGAGGAACCAGTTATTGCAT	183	58	M/wF	CT952199
032H24-SP6	F: TCAAAAGGGTAAGGTGCAGAG R: GCCAGAAAATGGCAATAGTTT	160	60	M	
032H24-T7	F: AAATGACAGCGTGTGGGAGT R: ACAGGACAGGTCCAGGTGAG	163	60	M/F	
039P6-SP6	F: GCAAAGGCTCTGAGAGAGGA R: GGCCACATGTCCTGTGTGTA	161	58	M	

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
039P6-T7	F: CATTTCTGATCCCCTCTTGG R: GCCAAGATGCTCTTGATGGT	155	58	M/F	
041O19-T7	F: ATTACATTTGCCCCAGACCA R: CTCTTGCTTCAGCGGTCTTT	249	58	M	
042.4B5-M13	F: TTGGGGAGACTTACCCACCT R: AGAGGCGTTAGGGTTGGTTT	100	60	M	
042.4B5-T7	F: TGGCTCTTTGTGTGGTGTAA R: TCCACAGACCCATGCAAATA	238	60	M	
047.1H3-M13	F: GGATCCTATGTCCTAGTTTGGAA R: TTTTTGTTCTTACCATTCATACAACA	113	58	M/wF	
047.1H3-T7	F: TTATGTCGACTTTGCCTGGA R: AGAATTCCATTAACAAGTTTTGGT	107	58	M/F	
047.2A3-M13	F: ACCCTTGTCACCTTCCAGTG R: CCTGCCCCTTATTCTTGTGA	194	58	M/wF	
047.2A3-T7	F: GCTGAGTCAACATCTCACATAGC R: GCCTTAAGATAACATCTGTGACCA	127	58	M/wF	
049.2F10-M13	F: GCACCAATACGCTAGAGTCCA R: GCTTGGCCATGTTAAGTGCT	176	61	M	
049.2F10-T7	F: GTTGGATCCTGCTGTGGACT R: ACAGCGTCCTTGATGCTTCT	212	60	M	
049.3F11-M13	F: TCAGCAGCAGCAGTAAGTGA R: ACGGTCAACCCATCTGTGAT	172	60	M	
049.3F11-T7	F: GCGAGTTCTGAGGACCAGAG R: GACCTGCCAACCAGTGATCT	173	61	M	
052H5-T7	F: GAACTCGCCTGTGGTTTCA R: GCCTGAATAAGATGCTGTCAAG	142	65	M	BV140796, CLY071

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
054A8-T7	F: GATGACCACGGGTTCTCAGT R: CCGTTCTCCTTCAGTCCAAG	181	64	M/F	
054F13-SP6	F: GCAGGGGCCTTAGTGATATG R: CATGTCTTGATCTGCCAGGA	189	58	M	
054F13-T7	F: TTGGGCTCAAGAAGTGGTTT R: TCTGCAGGTTCTTTGGATCA	164	58	M	
054J7-T7	F: GTGTGATCTGCTGTGCTTGG R: GCAGTTGCTGTGTGACTGTAGG	163	65	M	BV140803, CLY079
055N19-INT	F: TTTCAGCCATCCTTTCCAAC R: CGTGGATTGCTCCTTTGTTT	231	58	M	
060D8-T7	F: GACAGGAGGCACGTAAAGGA R: ATCTCCCCATCCCAAACTTC	171	61	M/F	
061.4H8-M13	F: CCACATTTTACACATGCCACA R: CCCAGAAAGACACCGTAACAA	117	62	M/F	
061G21-SP6	F: CAGGGATCGCAGACTCTAGG R: AGAGAGGGTGCAGAGCAAAA	227	58	M	
061G21-T7	F: TCTTCAGGAACGACAAACCA R: AAGGTACATTTTCCCCACTGAA	150	58	M	
061J18-SP6	F: TTGGTAGGATTTCCCATAGGTG R: TGATAGGTGAACAAACTTGAGGA	219	58	M	
063H12-SP6	F: ACCTCAGTAGGGGGCTTCTC R: GTGGAACAGGGTAAGGCAAA	155	58	M	
063I4-SP6	F: CAGCTACTTGTGGTCTGGTCA R: GGACTGACTGTAAGAGCCCACT	103	65	M	BV140776, CLY020
063I4-T7	F: TGTGACGGAGGCCAAAATTA R: ACTCCACATCAGGGTTGGTT	110	58	M	BV140774, CLY016

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
064P16-T7	F: TCAGAGATGGGCATTTTCAC R: CCTTATTCAAACAGCGTCCA	212	65	M	BV140805, CLY081
066M24-SP6	F: TTGGCTGCTCTCAGAATTGA R: GGATTGTGCATGTCATCTCG	169	58	M	CU001288
066M24-T7	F: ATGTGGGGACTGGTTCACAT R: GGGGAGCTTTCACAATCAAA	243	58	M	CU001289
067.1G8-M13	F: AGCTTTTTGGCCTTGTTCAG R: TGCTTTGGGAACTGACATTTT	206	58	M	
067.1G8-T7	F: CAAAAGCCACTGCAGGATTA R: CATGAATATGTGGGATTACATGG	182	58	M/wF	
069 E11-T7	F: CAGCCCTGGCATGTCTATTT R: GTTGCGAAGGGCTCTACAAG	237	60	M/F	
070F17-SP6	F: TTCCTTCAAGATCCCGTGAG R: CACAGTTGCTCCATTGGTCA	217	58	M	
070F17-T7	F: GCCAAATGTGACCCAAGAGA R: CCAGGAAGGAGCTAACACCA	102	62	M/F	
072G23-SP6	F: AGGTTTGGGGTTTTGTTTCC R: CTGATGGTCCAATGTCCTCA	163	58	M/wF	CU004599
072G23-T7	F: TTGTTCTGCTCCTTTCAGCA R: CTTGCACTTACGCAGTTCCA	162	58	M	CU004600
072G7-SP6	F: CCGTCTCAACTATATTTCAAAGTTTTT R: CGCAGGTCATCACTTCTCTG	194	58	M	
072G7-T7	F: CAACGCTTTCTTCTGACTCTG R: TTTTCCTGCCCTGTTTATCA	215	58	M	
077M19-T7	F: AGCCTGGGGTTGATATGG R: CTGTTCGAGATTCAGGTTGG	216	65	M	BV140801, CLY077

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
079.4H1-M13	F: AGTCCACACCACCACAGTGA R: TGATGGGAGTGGGAGTTTTC	248	60	M	
080.4F7-M13	F: TTTAACCACTCAGCCACAAGG R: CTTAACCACTGTGCCACCAA	245	60	M/F	
080.4F7-T7	F: ATCACTGTTGGTCGGCACTC R: TCTGCACATCAGCCTTGTTC	187	60	M	
081F8-SP6	F: GCAAAGGCTCTGAGAGAGGA R: CCATGCTTCTTGACCAGACA	182	58	M/wF	CU011557
081F8-T7	F: CAAACTACCACCTCCCAAGAA R: ACCAAAGGCACAAAGCAGTC	194	58	M	CU011558
083H5-T7	F: TGCCCTTATCTACGTTTTGG R: TCTGCAAAGCTGGATCTCTT	123	58	M	BV140789, CLY054
086.2F8-M13	F: GGTCCAGAATGCCTGAGTAA R: AGAGACCTTTTGTGGGTGGA	396	65	M	BV140826, CLY032
086.2F8-T7	F: CTGCCTATCTCCATTCTTCATAC R: AGTGTTTTGGGGCAAGTGTT	153	65	M	BV140830, CLY036
086J1-SP6	F: TCAACTTAGCGACTTCCTAGCC R: TTCAAAAATTCGATGTTGTCC	351	63	M	BV140795, CLY070
086J1-T7	F: GCCAGGATCGTGAGATATGG R: AACGTGCACGAGATAAGATGG	151	67	M	BV140794, CLY069
087.3A5-M13	F: CAAGGGAAATGGAGTCAAGG R: AAAGAAAGTGTGTGTGTCAGG	201	65	M	BV140828, CLY034
089.3B11-M13	F: GTCTTGTCCCCATCCTTTGA R: TGGCCCCATCTCTTATCAAC	158	60	M	
089.3B11-T7	F: CACAGCCTGAATGCAAAATG R: TATCTATGTTTTTGATGTTATTGACG	161	60	M	

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
090B11-SP6	F: TTGCTTGTGGTATTTCCTTGC R: CCTTTCTTCTTTCCTTACATGC	150	65	M	BV140791, CLY057
090B11-T7	F: AGTGGTTAAGGCACAATCCA R: CCCCTACATCAGTTCAACTTTTT	166	60	M	BV140788, CLY052
090G18INT	F: GGAGCAGATGTCAGCATTCA R: GCCTTTGGTGGCTGATTTTA	246	60	M	
090G18-T7	F: AAATCCAAACAGATATGACAAAGA R: GATTAAGTCTCTGCTGCCATC	155	60	M	
091.4G10-M13	F: AACTCCGGTAGGAGAGGAAG R:AAGGAGATGGACTCTTTTATTATTCC	152	54	M	
095.4B8-T7	F: TATAAGGGTACATGCAATTCTAACAAA R: CAAAGTCACGTGCTGCAGTT	150	60	M	
095.4F10-M13	F: GGGACAAAACCACTGAATGC R: GCTGTTGCTAGGTGGGAAGA	250	62	M	
095.4F10-T7	F: TTTTGCTTGCCACATCCATA R: CTCACGCAACTGAAAGTCCA	212	60	M/F	
097D2-SP6	F: TTTGGTTTCCCACTCTCTGTG R: CTTCCTTTGCAGGCTAGTGA	190	58	M	BV140810, CLY103
097D2-T7	F: GCAAGATTGGAAACATGAAGC R: TCACATTTCTTCTTTGCTCTATGC	177	65	M	BV140777, CLY021
100.3A11-M13	F: AATTCGTCTGCCACTGGAAC R: AAAAAGCGACAATGGAGTCG	217	55	M	
100.3A11-T7	F: TCCCCTGACCACTAAGGTTG R: CTTCACATGGCGTGAAAATG	197	58	M	
100.4F5-M13	F: AGAGGATCCGTCACGATTTG R: TGGAGCCTTCTGAGGTTAGC	162	58	M	

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
100.4F5-T7	F: TGCAAAATTCAAAGGTCAGC R: GGCAACAGATGCCTGGTATT	152	58	M	
100H13-SP6	F: GGTATTTGCCCAGTGACCAG R: TCCCTGCAGTGCATTGTAAA	165	62	M/F	AJ542956
101H8-SP6	F: GTTGTGCCCTAAGGCAGGTA R: TTGGCAAACTATGCCGAGTA	170	58	M	
102J15-SP6	F: CCACAGTGCATACAGCAACC R: GCTGATCTGCTTGTTTTCTGG	190	58	M/wF	
103.3A6-M13	F: GTGGAACTGCGCACTGCTTA R: CAGGAACATTAGGCCTCAGC	150	65	M	BV140840, CLY058
106F1-SP6	F: ATCTGAAGGTGGAGGTGCTG R: TCTGCTCTCCTGGGGTTCTA	244	58	M	
106F1-T7	F: CTCCTGATGACCATGGGACT R: TGCCCTAATGTCTGGCTTTT	153	58	M	
106J17-SP6	F: ATGCCTTCTACCACCACTGC R: ACCACTGCACTGAGCATGAC	161	58	M/F	
106J17-T7	F: TGGCATCTCTCATTCACCAA R: TGTAAGCAACCAGTGGAACG	158	58	M	
107.3H9-T7	F: GTCAGCACGACAGCTCACAG R: ACAGCAATAGTCCACCAGCA	247	68	M	
108.4C7-M13	F: AGCGAGGTCTGCACTTTCC R: GCGGGGAGTACATCAGTTCC	156	66	M	BV140827, CLY033
108.4C7-T7	F: CTACGTGAAGGAATGTGTCTGG R: GATGTTTCTAAAGTCCAGCAAGG	150	60	M	BV140832, CLY038
110.3H12-M13	F: GGGCCAGAATATGCAAGGA R: GATGTGTTTGTGTGCCTCTGTT	182	60	M	BV140834, CLY047

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
110.3H12-T7	F: GCCTCAAGTAGAACCACATCC R: GCATCCAGAACAGCAAACC	300	60	M	BV140837, CLY050
111.2F5-M13	F: TGGAGAATTCACTGGCTGTC R: AAATGAAATAGCTGCAATGAAGT	102	60	M	BV140842, CLY060
111.2F5T7	F: GGCAGGAATCCCACACATAA R: CAGGAACATTAGGCCTCAGC	116	60	M	
112.1A9-M13	F: AACAGAACCACTGCACTAAACC R: CAGATTCCCTTGGCTGACC	358	60	M	BV140835, CLY048
112.1A9-T7	F: TGTCAGCTTTGCCATTGTCT R: TCTGCCTGAAATGAAAGGAA	226	62	M	BV140836, CLY049
112E12-SP6	F: CTCCTTAGGGTCTGCAGTGG R: TTGTGGACAGGCCTGGTAAT	216	58	M/wF	CR955686
112E12-T7	F: AGTTGGGACCCTCAACTGC R: AGGCATGCATCATCTCACAG	177	58	M/F	CR957122
114E24-T7	F: GTGTGCACTAGGCACCCTCT R: TCAGGTGCTTGCCTCATGTA	240	58	M	CT006811
114I17-SP6	F: CCCCATGGTGTTATTCATGT R: GTGGCATTTCCAAAAATCAG	172	58	M	
114I17-T7	F: GAGCCACAGCACATTTCTCT R: CAGGTTATCCCGTGACAAAG	215	58	M	
117.4F7-M13	F: TGGGTTTCTTGGAATCCTGA R: CCAGTGAAGGGGGACATCA	248	55	M	
117.4F7-T7	F: ATATGCCAGACATGGCACTG R: GCCCTGATCTAACTACTGCCAATC	148	60	M	
118.1A9-M13	F: TGAAAACAAAAACCGAACAGG R: ATGGCATGCAAACACAAAAA	157	58	M	

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
118L7-SP6	F: GGGCTACAGGAGGACATGAG R: GCACAAGTTTGGCCGATT	177	60	M	
118L7-T7	F: ATCTGCTCCCCTTTGGTTTT R: CCCCAGATTTACTGCCTTTG	225	55	M	
118N21-SP6	F: GAGAGGGAGAGAGGTGGACA R: GGGAATGTTCTATTGCTCCAA	275	58	M/F	CU025698
118N21-T7	F: GCAAATAAAAGTGGGGGACA R: TCATGTGGGAACCAGAGACA	153	58	M/F	CU025699
119K22-SP6	F: CTGGGTCTAGCCCCTTATCC R: ATCGAGGGCTTGAAGCTGTA	158	50	M	
120.1A5-M13	F: ACAGTGCGTTCTGTGGTGAT R: AGTGAGCTGAGAATGCTTTGG	174	60	M	BV140838, CLY053
120.1A5-T7	F: GAACCAGCACTGCTCATCAA R: CCTCCAGAATGTCTCCTCCA	210	60	M	BV140839, CLY055
120A19-SP6	F: CTGGCTGGAGTGCGATCT R: ACAACCCGTCGGATTCTCC	151	55	M	
120A19-T7	F: GACTAACCACGTTGGGGAGA R: TATACATGTGCTGGGCTTGG	151	58	M	
121G24-SP6	F: CCTTGAGTCATTGCCCTCAT R: GTTGGTGAAGCACATGTTGG	246	58	M/wF	
121H9-T7	F: TCTGAGACCTTGCGAATCCT R: TGAATCCTTTCCCAGTTCCA	166	58	M	
124.3G9-T7	F: TTACCAAGTCGCTGTGGTCTT R: GATTCTGCATTTCTTGAGCTCTT	151	58	M	
125.3G11-T7	F: TGTGCATGAAATCTGGACTCA R: TGCTGTGTAGGAGGCATTTG	164	65	M/F	

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
125H6INT	F: CAAAAGGAACAACCCCACTC R: TTCACATGACAGATTCAGCAAA	242	55	M	
125H6-T7	F: CCAGCACCCTACTTTGAGGA R: AGGGCATGGGAACTTTAACC	204	58	M/wF	
126G2-SP6	F: TCAGGCAGGAGAGCTGATTT R: CAGATGGGTGTGTCCTCCTT	249	58	M	
126G2-T7	F: GCAACTTGCACTGATTGTCC R: ATTTGTGTGGAGGGCAGGT	200	58	M	
129K23-SP6	F: ACACACATACCTGGCTGTCC R: GGTGAGCAGATCCCACTTCC	112	67	M	BV140875, CLY076
129K23-T7	F: GGGGAAAACTCCAAAGCAG R: CAGGGGAACAAAGCCAGAG	133	65	M	BV140800, CLY075
131N23-SP6	F: GCTTGTCCTAATTTGCCTCTG R: CTTTGAGGACGGGTAAATTG	152	58	M	
132K10-SP6	F: AGGGGCATCTTCACTCACTC R: ATGTCACAAGGTCCCTGAGC	220	58	M	CU035255
132N15-SP6	F: GAGCCACAGCACATTTCTCT R: CAGGTTATCCCGTGACAAAG	214	58	M	
132N15-T7	F: TGTCACAGCCATCTTTCAGA R: GTCTGCAGGCTCTCATGATT	186	58	M	
134H14-SP6	F: AAGGGAAAGCAGGTTCCAGT R: ACAAGCGGTCTGAGGACACT	191	60	M	
134I16-SP6	F: AGTTGAAGCAAGTGTGTGTGG R: CGGACTACGCCAAGAAAAGG	188	62	M	
134I16-T7	F: CCATAGAGTGAGAGCTGATTG R: TTGTTTTCCTTTTGGACTGG	238	62	M	

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
137I17-SP6	F: AAGGCCAAAGACATTCTCCA R: ACCTCCTGAGTTGGCTGAGA	213	62	M	CU038728
139C20-SP6	F: GTGGGTCTTCTTGTGGCAGT R: ACAAGAGAGGGCCCTGGAC	109	58	M	
142O2-T7	F: AATATCCACAACCCCTCTTCC R: CGGAGCATAGTAGCCAGACC	155	58	M	
144B9-T7	F: TCAATGGCCCAAAGGTCTAC R: GGGGAGCACCAGATATGAGA	156	58	M	
145I6-SP6	F: AGGCAAGTTGTAGGAAGAGCA R: TTTAAAAGGATCATTTCATTCCA	164	58	M	CU046684
145I6-T7	F: AACGGCCCAGGCTAGAGTAT R: AGAGGTGGTTTGCCTTGCTA	213	58	M	CU046685
148G3-T7	F: CAGAGGAGCGTCTTCCAGTT R: GGGTTTTTCCCCCAGTTTT	159	60	M	BV140785, CLY045
149H18-SP6	F: GCGGTCTCAGTTTCTCTTCC R: CTGCTTCTTTCGCCTCTCC	242	65	M	BV140784, CLY043
152E2-SP6	F: CGGTGAATTTTGAGGAAGGA R: TGTGCCTTTGTCAGGAATCT	174	TD60-50	M/F	
152E2-T7	F: CAGCATGAGAAAGAACTATAATAGGC R: AACAAAACATGGTCGATGTAGAGA	150	58	M	
152G20-SP6	F: GATGAGGGGCTATTCAGGTT R: ATCCCACACGTGAACTGTCT	246	58	M	
152G20-T7	F: TTACTCCAGGTAACGCTTGC R: GAAAAGCCTGTCCTCCTCTC	182	58	M	
155B8INT	F: TAGCACCCAGGAGGTGTAGG R: CCTGCGACTAAATTGTTCTTCC	165	60	M	

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
155B8-T7	F: GGTACCAAAGGCACCAGAGA R: CCTCCTAAAGATGGCATGGA	225	60	M	
155M11-T7	F: GGGGAGAGGATCTTCATAGGG R: TCCAGGCTGCCATGTTTTA	247	50	M/wF	
159E3-SP6	F: TGAGCCCCCACAGTAATTTT R: TGAAGCCAGCCAGAGTTTTT	237	58	M	
159E3-T7	F: TGGTAAGCAACTAAGATGAAAAGG R: CATCAACACAAAAGTTAGACAGGAA	246	58	M	
159F5-SP6	F: ATGCGTTAGCCCACTCAAAG R: TTTTTGCTGTGGTCTCTGGA	241	58	M	
159F5-T7	F: CCACAGCAATGACAAACTGC R: AGTGTTGGGCATCAGGTTTC	169	58	M	
160K10-SP6	F: TCACGTCCCTCAACAAACAC R: GGGACACGGTAATGAGGAAA	198	60	M	
160K10-T7	F: CAGCTCAGGGAAGAGACAGG R: CATTTCCACACGGAAGTCCT	247	60	M/wF	
165 E24-SP6	F: ACCCAGCCAAAAGCAGATAC R: CTGAGGCCAGGTGTGGAG	171	62	M	
167N20-SP6	F: CCATGGAAAGTGCAATGGTT R: GCATCTAGAAAAATGGAAGGTGA	214	58	M/wF	CU062306
167N20-T7	F: CTCCAAACCTCCACTTCCAG R: GAGCTGCTGGTCGATTTTTC	187	58	M/wF	CU062307
168I4-SP6	F: TCAGAAGGGGAAGAGAGTGG R: AGGCTCGGCTCGATTTTT	138	55	M	
168I4-T7	F: CCATTTCAACTCCTGATCCA R: AAGGGTAGGTCAGAGAGGAAGG	101	65	M	BV140781, CLY031

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
168O8INT-2	F: AAGTTCCTGAAATCCCACTCC R: CATCTGAACCCCACTTCACC	200	58	M	
168O8INT-3	F: ATACCTGGGAGAGGAATTGG R: CAAGGAAAGGACCGTGTAGG	310	58	M	
172I8-T7	F: GCCCTTTCCTTCTCTCTCAC R: GCCACCACACGAGAATAAAC	196	60	M	
179K8-T7	F: TCATAAGCCTCAGTGGAAGC R: CCCTTTTACTTCTCCCCAAA	200	58	M	
180P20-SP6	F: CCCGTCCTTGTACTTTGGAG R: ATTTCACATGGCCCCTAATC	128	60	M	BV140790, CLY056
180P20-T7	F: ATAAATGCCGGAATCCATGC R: CCCCATGGGAATGGTAAAGT	181	52	M	BV140787, CLY051
181B18-SP6	F: CCCGCCAAGTCTATTTCC R: CAGTTAGTGGGAGGTGAGACG	156	60	M	
185M14-SP6	F: TCAAGACTCTATCACAGCACTAAACAG R: AAGGGTCCACCTCAGTCACA	100	50	M	
185M14-T7	F: AACCCATTTTCTCACAGTCTTG R: CAACACAAGTTGGAATGAGATG	106	60	M	BV140782, CLY039
186J13-T7	F: AATCATAAACAGTCCATGGTCA R: TCAGTTCTCCCACCTGGATA	151	58	M	
188E20-SP6	F: TCAGGTGAAGATTAAAGGAAGC R: GGGAATCCAACCAATAAGGAA	250	58	M	BV140780, CLY030
188E20-T7	F: CTCAGTGTCAGCAGGTTCCA R: TGCAGCTCTCTATCAGAACAGG	185	58	M	BV140778, CLY023
190M2-SP6	F: CTGATGCAGGTTTCACTGG R: GAGTCAGAGAGCTGGAACTGG	109	67	M	BV140775, CLY018

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
190M2-T7	F: CAGCATGGCTTCACTATTTCC R: TTTAGACAGGCAGAATCAGTTCC	164	66	M	BV140874, CLY015
205D10-T7	F: GGGCTATGGCCTAGGAGAAC R: CTAGCAAAACATCTTGTGACAGTA	169	62	M	
207D10-T7	F: AAGGGAAAGCAGGTTCCAGT R: ACAAGCGGTCTGAGGACACT	191	58	M	
207P5-T7	F: CTGGAGTTCCATTGGGAGGT R: CACTCACAGGCAGCACATTT	169	58	M/F	CU102185
209K10-SP6	F: TGGCTTCCTGCTGAAAGATT R: ATGGCAGCAAGTTTCTCTGG	203	62	M	
215C6-SP6	F: AGCTTGGGTTCATGGTCAAA R: CCTCTCCCTATGGATCATGC	221	58	M	CU108088
244B13-T7	F: GCTGGACAAATCGGTGTCTT R: CCTTGTGGACTATGCCACCT	171	62	M	CU126281
263G23-SP6	F: GTGCCTCTGGGCATATCTTT R: TAATGTCAGCGGAGATGCAG	150	58	M/F	CU140369
263G23-T7	F: GATGACCACGGGTTCTCAGT R: GGCCCAGGGATTAATTTTGT	168	58	M	CU139968
269J9-SP6	F: CCGTGGCAACAAATGTTAGA R: CCAGAGAATGCTCCTGAACC	191	58	M	CU143890
269J9-T7	F: CCATGGCAATATCTCCCTTT R: CCTGCAAATTGTGTGAAAGC	184	58	M	CU143373
275P16-SP6	F: GGATCCTTACCACAGCCAGA R: CACATGAGGATATGAGGAGATACG	249	58	M	
275P16-T7	F: CGACTCACTATAGGGGAGAGGA R: GGTGCTCTTCGTTACCTGGA	157	58	M	

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
278M12-SP6	F: GTTGGCAGACGTAAGAAATGC R: TGAAAATGCTGGTTTTGTGG	244	58	M	
278M12-T7	F: TATTACGGCCGTTACCAAGC R: GTGCCTGGATAGGCACATTT	189	55	M	
280P20-SP6	F: CACAGTCATTTCTGGGATGC R: CCTGAGACTATGGGCAGTGA	242	58	M	
291D19-T7	F: TGAGTTTGTGGTTAAAATGACAGG R: GGAGACAAATAGATGGACTCAGAA	115	62	M	
318M1-T7	F: AAGTCAGGGACCAGCAAAGA R: TCCCTCTGAGAACCCAAATG	193	58	M	
324H11-SP6	F: GACAGGACAGCGGAGTAAGC R: GTCCAGCAATGCACTCAGAA	191	60	M	
324H11-T7	F: GCTGCCTGAAACCTGACTTC R: CATACACATGTCCCCTGCAA	217	60	M	
329G16-SP6	F: CTCACCACTTGTTCTGGAAGC R: GGTGGAACTACTTGCTCATGG	204	60	M	
331E10-SP6	F: TCATTAGGTCCTGGGCTCAC R: ATTGTTGGATGCTGGGACTC	172	58	M/wF	
331E10-T7	F: CAGTTTCCCCACACTGAACC R: TCCATGGAGCATAGTGGATTC	105	58	M/wF	
335P13-SP6	F: AAATGCCAGACACCTTCCAG R: TTCTCCTGTTTTTCCTTTTTCG	210	58	M	
338A6-SP6	F: TCAACCCTGAATCCAGTCTCT R: TGCCTACTCTTCCCCCTAAAA	249	62	M	
338A6-T7	F: CTGGGGATTTTGCTTGATGT R: CATGGGGAAATTCTATTTTAGGC	154	58	M	

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
341G20-T7	F: GCAAGGAATGATGCACAAGA R: ATTCCTTTGGTGTGCCTTTG	178	58	M	
344A12-SP6	F: AACCTTGCCTACGGGCTTAT R: ATGTTTGTCCCCTTGTCAGG	152	58	M	CU312820
377O23-SP6	F: GAATGCACATGTAGTGCCGTA R: TGACTGATGCAGCAGGAAAG	151	58	M	
394K12-SP6	F: TTCCATATGCCAAGGAAGC R: TTGTCTAGGACAATGTTTTTCAGG	151	60	M	
394K12-T7	F: GGGAATCTTGAGTCGCTTTG R: GCCTTACTCTGAGGCTGTGG	249	60	M	
395L19-SP6	F:ACCCTGAATGCCAGTCTTTG R: CTGCCCAGTTTTCTCCTCAG	170	58	M	
417N24-SP6	F: CCCTAGGGCAGGTAAACAGA R: TTGGCAAACTATGCCGAGTA	164	58	M	
417N24-T7	F: GCAATGCAAACGAGAACAGA R: AACAGCCAACCTCAGAGGAA	239	58	M	
418J18-SP6	F: TCTGCTGGCACATTTGTTCT R: CCAAAAGTTCAGGCAGCAAT	250	58	M	
418J18-T7	F: CTCCCCCTCCAAACTACACCA R: GGGAGGGAGCAAAGTTCTCT	163	58	M	
422E23-SP6	F: GGCAGCATTTTGGTAGAAGC R: ATAGCGGGGTGGATGTACTG	177	58	M	
422E23-T7	F: ACGCAAAAGGTTCATTACCG R: TTTCCATCTGTCATCCAATCAG	175	55	M	
437I11-SP6	F: GCAATGCAAACGAGAACAGA R: AACAGCCAACCTCAGAGGAA	239	60	M	

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
437I11-T7	F: GAGGATTTCCCTTCCCACAT R: TGCGTCCAATATGTCTACCC	241	60	M	
450C22-T7	F:CCTGTTACAGCTCCGCTTTT R:CTGGTTGCCCAATTTGTTCT	151	58	M	
504H13-SP6	F: CCCTTCCGCACTGTAAATGT R: TCCTGTTACCCACACAACCA	163	58	M	
510F11-SP6	F: GGCAAACTATACCTAGCAC R: TCCGCAGTCTGATAGAGCAA	213	58	M	
510F11-T7	F: AATCTGGGGAAGCTGTGCTA R: TTGATGGTGGCAGAATGAGA	155	58	M	
ABW-M13	F: GCGAGGGGTGAGTTATTGAA R: TTAGCTGGCCCTTCCCTTAT	193	60	M	BV140869, CLY088
ABW-T7	F: TATGGAACCCCTTCTGCAAG R: CCTGCCGACATGACAAAATA	209	65	M	BV140868, CLY087
BBW-M13	F: CAGTCCTGTCAGTGCTCCAA R: CATGGCACAATGCAACTAGG	157	60	M	CLY063, CLY063
BBW-T7	F: CCCACCTCAGGATATTGCAT R: GCAGTGGTGTACAAAGACAGCA	214	65	M	BV140862, CLY061
CBW-M13	F: GAGGAGAAGTGCCTTAAATTCC R: CCCAGAAAATCGTTCATTCC	177	60	M	BV140870, CLY089
DBW-M13	F: GTGCCTGGGGATTCTCAGAT R: GTATAGTTTGGAAGGTGCTAAACACAG	150	60	M	BV140866, CLY065
DBW-T7	F: CTGCTGAGCAAGGGGTTAAG R: CGAGTGTTTCAGCAAACAGG	195	65	M	BV140865, CLY064
EBW-M13	F: TTAGAATGGGCTTGGCTCCT R: ACAGCTTTGAGGGATGGTTC	179	60	M	BV140858, CLY025

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
EBW-T7	F: TTGAGCCCTTGTTTCTGTCC R: CAGGGAAGTAGGGAGTGACC	165	60	M	BV140859, CLY027
FBW-T7	F: AGCTGATCGAACCCATAACC R: CTCTCCCAATGCCCTTCC	173	60	M	BV140861, CLY044
GBW-T7	F: ACTTTGCACTTGCCCCATAA R: TCCCAGCTAAAAAGGTACTCC	120	60	M	BV140860, CLY040
HBW-M13	F: GTTGCTCTGGGTTTTCTTCC R: ACTCACCATCACACGACACC	247	60	M	BV140872, CLY093
HBW-T7	F: CGTACTGCTTTTAGCTTGAGACC R: AAGAGGTGAGTCCGTTGACC	236	60	M	BV140871, CLY092
JBW-M13	F: CCTCTGCTGGTCACCTTCTG R: TTCCACATAGAAGCCCCCTA	242	60	M	BV140873, CLY094
MBW-M13	F: AGAAGGAGCCTGAGACAAGG R: CTCAACACCCAGTCTTTTTGG	154	60	M	
OBW-M13	F: TGTGGCAGAGTGAGCTTTCC R: CATTCCAGTAGAGGGCTTCC	348	60	M	BV140856, CLY017
OBW-T7	F: CCAAAACTAATGAGAGCCGTGA R: CACAGACTCAGGAGCAGGAA	173	60	M	BV140857, CLY019
SH2A1	F: CGGTGTCAGGTTTTGGACTT R: AAGGATTCTGCTGCCCTCAT	747	64	M	BV005744
SH3B14	F: GTGACCTCCCAGGAGCTGT R: TCTGCCTATGCTCTGGTGAA	486	64	M	BV005745
SH3B19	F: AAGCCTTTCATGGAAATTGG R: TTACGCAGACATCCTGGACA	255	58	M	BV005720, Y3B19
SH3B6	F: AGAGTGCGATTTGTGATGG R: AGAGTCAGAAGAAAGCGTTGAT	492	64	M	BV005718

Table A2 continued

STS marker	Primers 5′ to 3′	Size, bp	Ta °C	Male or female amplification	GenBank Accession and alias symbol
SH3B8	F: CCCAAGTTCCTTGCCATC R: AAATTGAAGAGGCCCCAAAG	472	58	M	G72337, Y3B8
Y2B17	F: TTCAGTCCTGCTTTCTCCTCA R: CAGGATGTGCCATGTGATTG	528	58	M	G72335
Y3B1	F: TGGGTTAATGGGATTTGGTG R: CAAGCACAGCTCTGTATCAA	508	58	M	G72336
Y3B12	F: GGGAGGCACTGGAAAGTACA R: GGTGGAGGAATCAGCTGGAG	400	58	M	G72338
YA16	F: TGACTGGAAATTGAAGATG R: TTGTAGCAACAAAGTAACAC	157	58	M	BV005729
YE1	F: CTTCACTCCCGACCAAGAGA R: GTGTGTCGTGCCGTGTTTAC	199	58	M	BV005727
YH12	F: CGAACAGGTGACGAAGCATC R: GCAGACATGCACACCAACC	98	58	M	BV005747
YJ10	F: AGTTCCCCTGCACACCT R: TGCCTCCCACAGCCATAC	215	64	M	BV005728
YM2	F: TGGTTCAGATGGTGTATTTTGTT R: TTTGCAGCCAGTACCTACCTT	119	58	M	BV005725
YP9	F: AAGCACTGCCTTTTGGAATC R: AACCCTGGACTTTCTTTGAA	216	60	M	BV005726

(M: male, F: female, M/F: male and female; M/wF: male and weak female)

Table A3: Summary information about the MSY contig map.

Contigs		Genes	STSs	Number of BACs	Average BES size, bp	GC% of BES	Number and % of repetitive BES
Contig I	Heterochromatin	3	3	2	557	46.2	2 (50%)
	Single copy	2	33	28	756	42.7	8 (14.3%)
	Multicopy	13	63	52	623	42.6	41(39.2%)
Contig II		11	87	55	777	42.1	31(28.1%)
Contig III		2	47	27	660	38.5	9(16.7%)
Contig IV		2	31	20	683	38.0	10 (25%)
Contig V		3	18	13	856	31.1	8 (30.7%)
Total/Average		**36**	**282**	**197**	**701**	**40.1**	**109 (29.1%)**

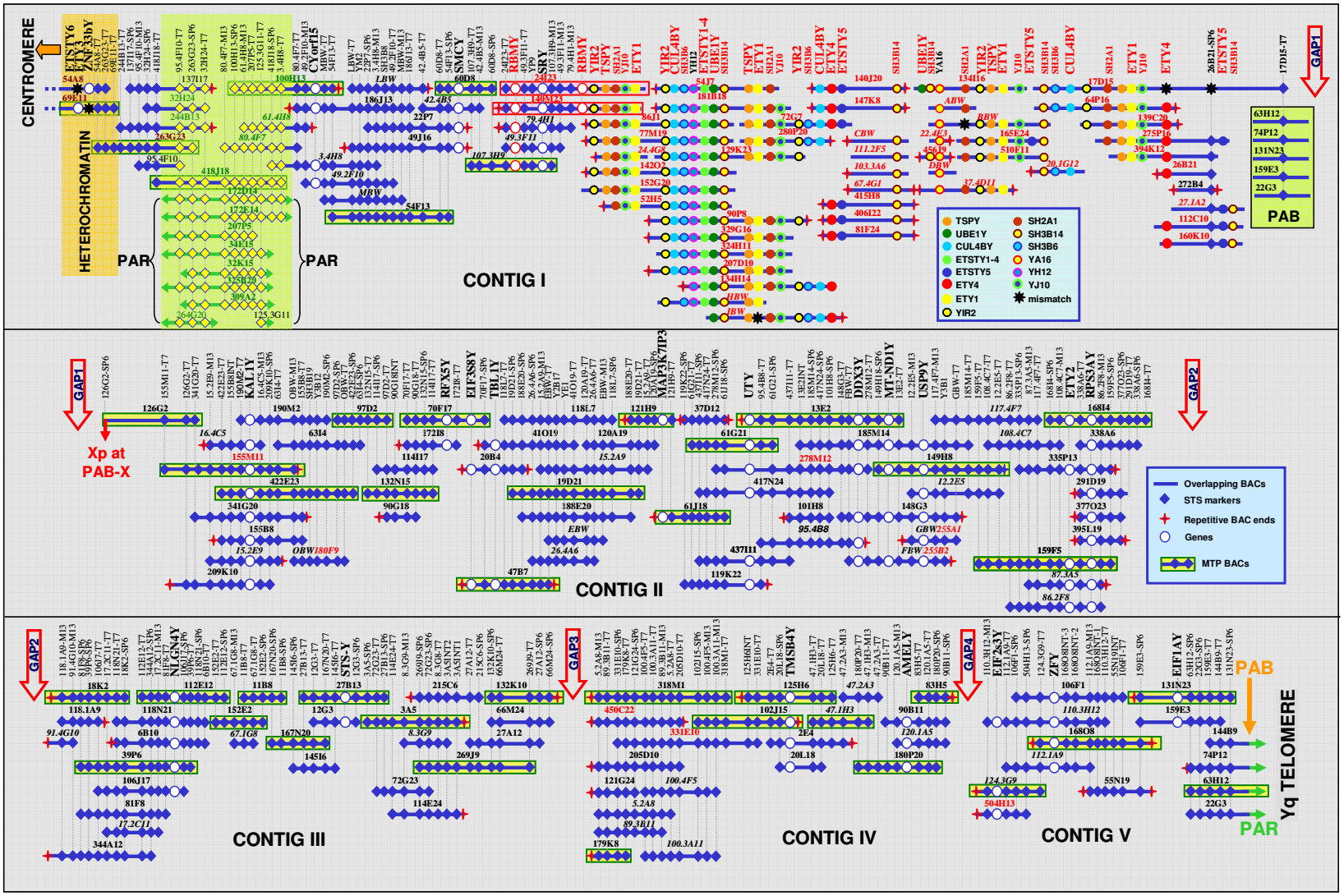

Figure A1: Comprehensive map of the male specific region of the horse Y chromosome.

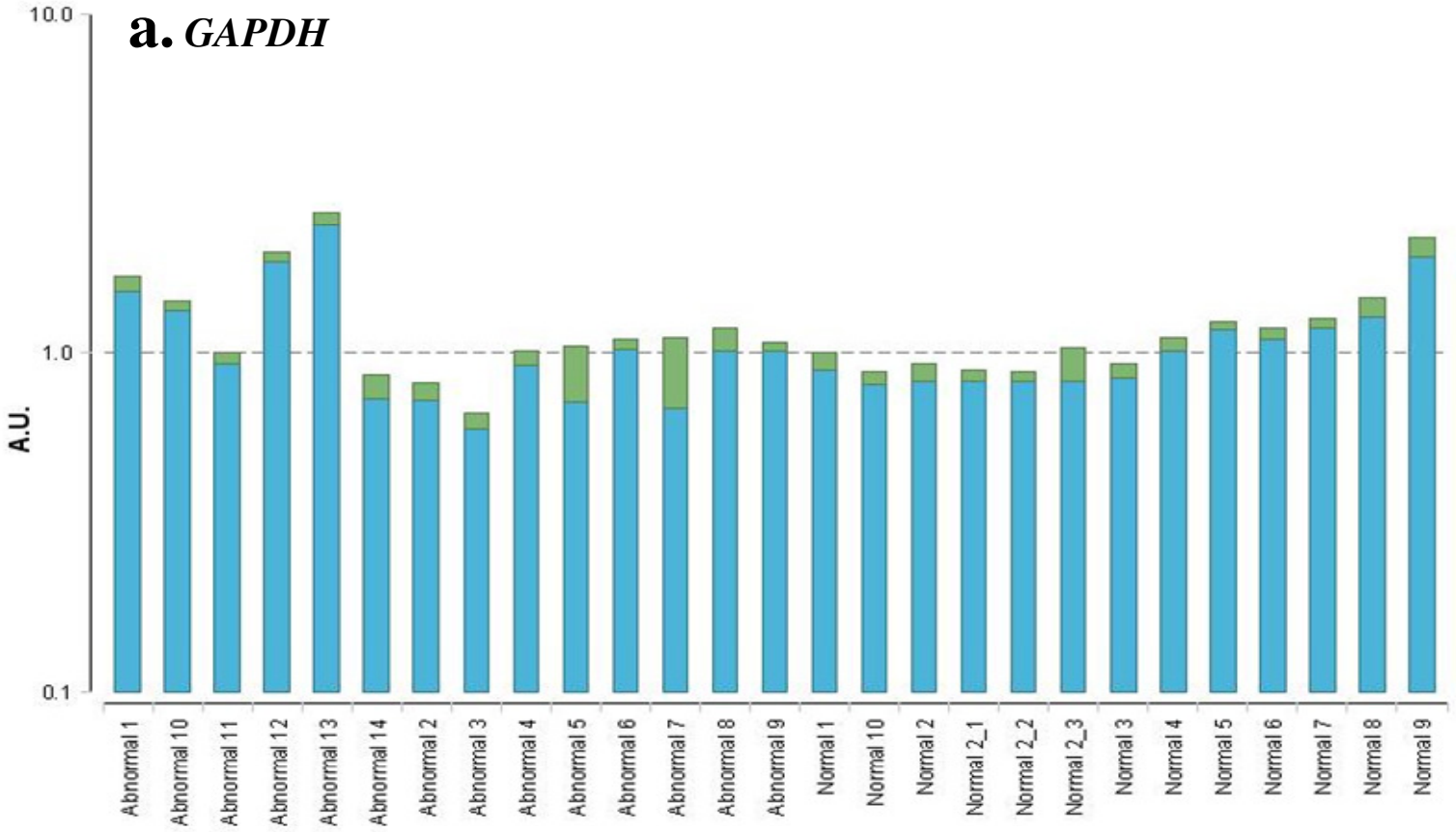

Figure A2: Expression analysis of ECAY genes and reference genes in normal fertile and infertile/subfertile (abnormal) stallions (represented data obtained directly from qBase software) **a.** Expression of *GAPDH* (reference gene): showed uniform expression across fertile and subfertile group of animals.

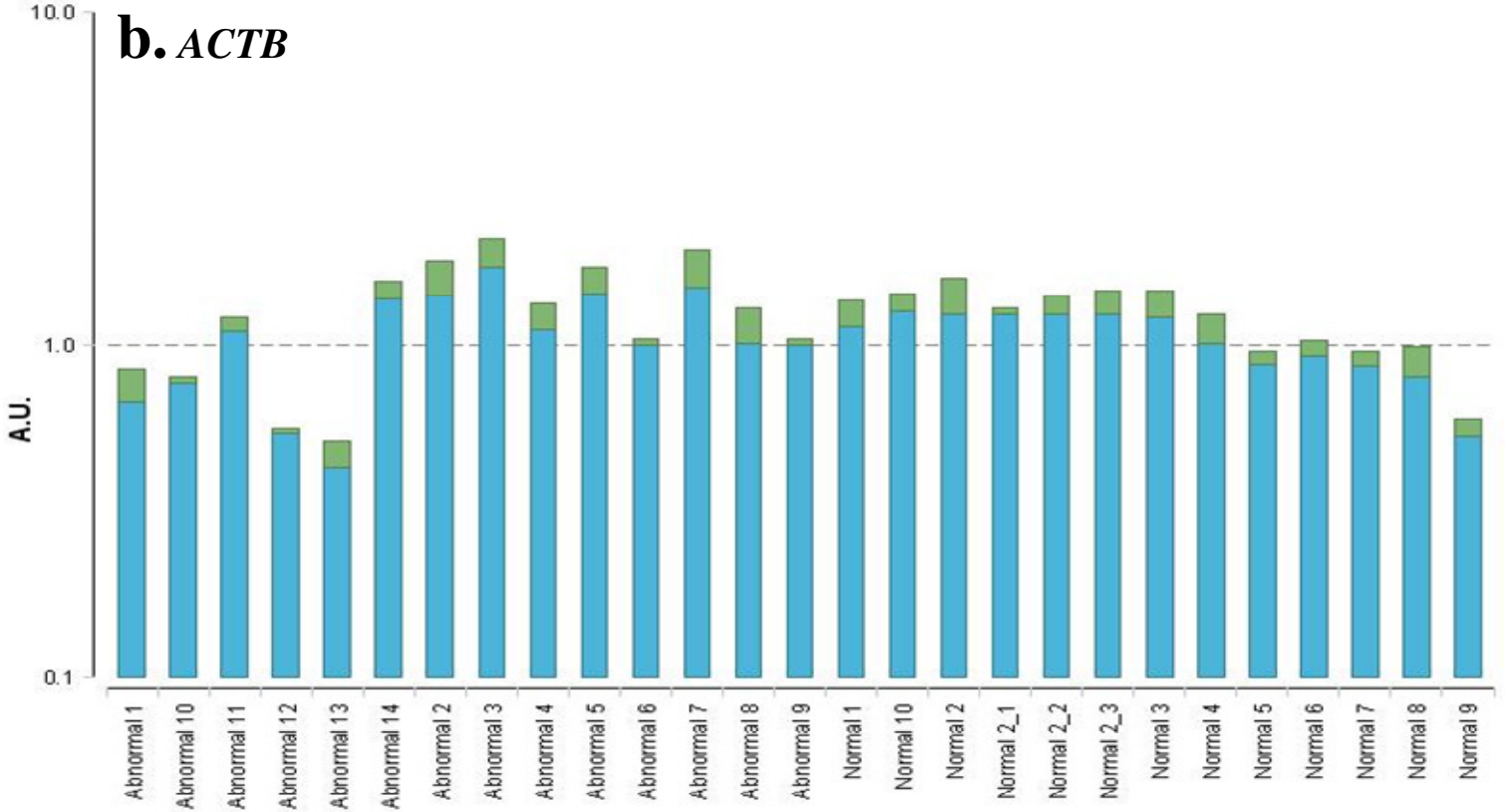

Figure A2: Expression analysis of ECAY genes and reference genes in normal fertile and infertile/subfertile (abnormal) stallions (represented data obtained directly from qBase software) (contd.) **b.** Expression of *ACTB* (reference gene): showed uniform expression across fertile and subfertile group of animals.

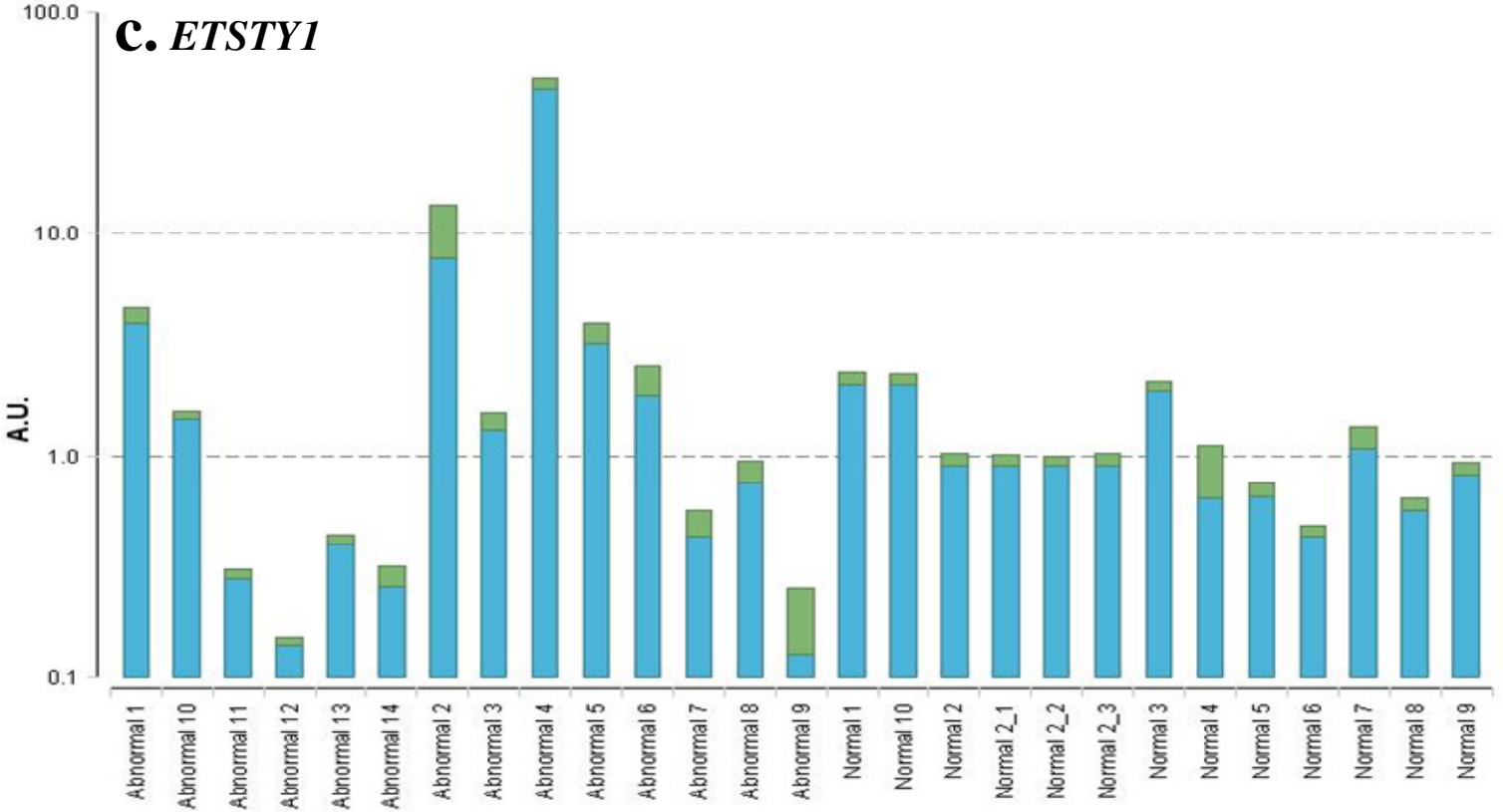

Figure A2: Expression analysis of ECAY genes and reference genes in normal fertile and infertile/subfertile (abnormal) stallions (represented data obtained directly from qBase software) (contd.) **c.** Expression of *ETSTY1* (multicopy, testis-specific): showed differential expression in subfertile/infertile (abnormal) group compared to normal, fertile group of stallions.

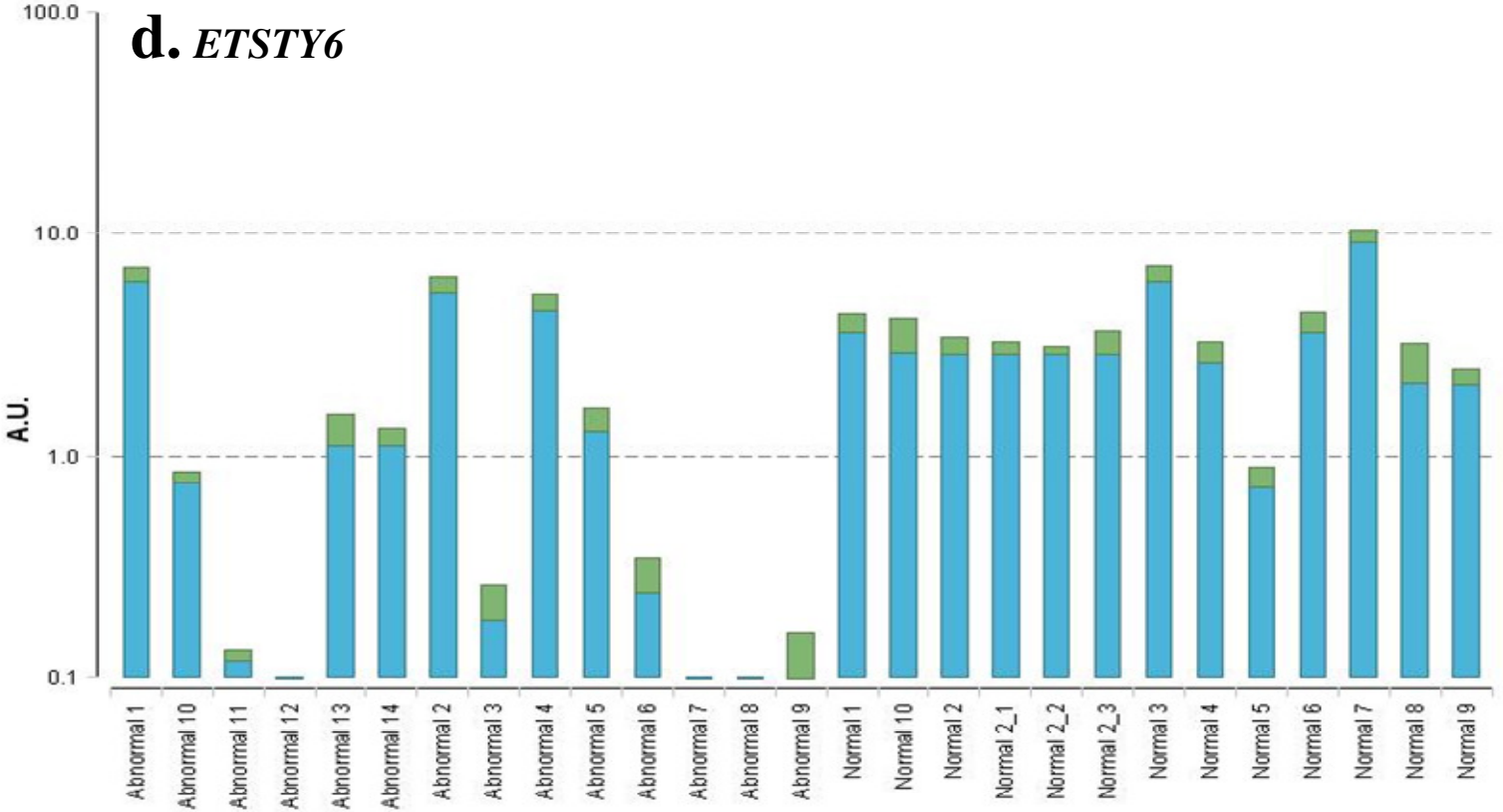

Figure A2: Expression analysis of ECAY genes and reference genes in normal fertile and infertile/subfertile (abnormal) stallions (represented data obtained directly from qBase software) (contd.) **d.** Expression of *ETSTY6* (multicopy, tetis-specific): showed differential expression in subfertile/infertile (abnormal) group compared to normal, fertile group of stallions.

VITA

Name: Nandina Paria

Address: c/o Dr. Terje Raudsepp, Assistant Professor, Veterinary Integrative Biosciences, College of Veterinary Medicine, Texas A&M University College Station, TX-77843-4458. Phone: 1-979-862 2879

Email Address: nandina1@gmail.com

Education: B.S., Botany (Hons.), Zoology and Chemistry, University of Calcutta, India, 2000
M.S., Biophysics and Molecular Biology, University of Calcutta, India, 2002
Ph.D., Biomedical Sciences, Texas A&M University, USA, 2009

Awards: Best Graduate Student Platform Presentation Award, Texas Genetics Society, 2008
Fisher Institute Medical Research Award for research in Cell Biology and Genetics, College of Veterinary Medicine and Biomedical Sciences Texas A&M University, 2009
Session Winner and 2nd place in Genetics for oral presentation, Student Research Week, Texas A&M University, 2008

Publications and Manuscripts: Mondal BC, **Paria N**, Majumdar S, Chandra S, Mukhopadhyay A, Chaudhuri U, Dasgupta UB. (2005) Glutathione S-transferase M1 and T1 null genotype frequency in chronic myeloid leukaemia. European Journal of Cancer Prevention 14, 281-284.

Chowdhary, B.P., **Paria, N.**, Raudsepp, T. 2008. Potential applications of equine genomics in dissecting diseases and fertility. Animal Reproduction Science, 107, 208–218**.**

Nandina Paria, Allison J. Pearks Wilkerson, Charles C. Love, Dickson D. Varner, Patricia C. O'Brien, Malcolm A. Ferguson-Smith, William J. Murphy, Bhanu P. Chowdhary, Terje Raudsepp. A gene catalogue of the horse Y chromosome (manuscript in process).

Terje Raudsepp, **Nandina Paria,** Keith Durkin, Bhanu P. Chowdhary. Sequencing-ready clone map of the male specific region of the horse Y chromosome (manuscript in process).